**PHD**

# Population analyses of bacteria / host interactions

Fleming, Vicki Murray

*Award date:*
2006

*Awarding institution:*
University of Bath

# POPULATION ANALYSES OF BACTERIA / HOST INTERACTIONS

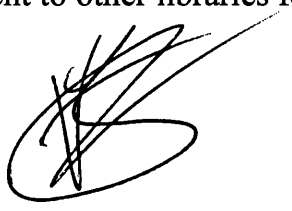
Vicki Murray Fleming

A thesis submitted for the degree of Ph.D.

University of Bath
Department of Biology & Biochemistry

October 2006

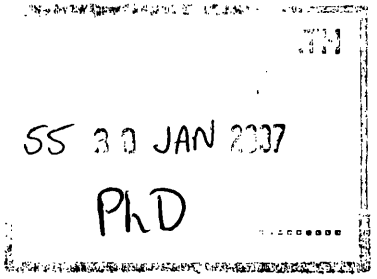








ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346


30 JAN 2007
PhD

I hereby declare that this thesis is my own work. Any contribution made to the research by others, with whom I have worked at the University of Bath or elsewhere is explicitly acknowledged in the thesis.

## ABSTRACT

This thesis consists of several studies investigating the evolutionary history and population structure of bacterial host interactions in three genera; two natural bacterial populations, *Wolbachia* spp., *Vibrio* spp, and an experimental bacterial population, *Staphylococcus aureus*. The studies presented are firstly introduced with background to the factors which drive microbial biodiversity, the lack of an appropriate bacterial species concept and the ramifications of this, and the current methods available to examine bacterial population structure. Understanding the extent of local adaptation on a global scale is crucial for development of a proper understanding of the genetic structure of any bacterial population. With a view to this, *Wolbachia* spp. is firstly investigated in chapters three and four using multi-locus sequence analysis to identify levels of host congruence, adaptation and horizontal gene transfer within i) a global and ii) a local sedentary *Wolbachia* population. Chapters five and six address the question of adaptation on both a local and global scale within *Vibrio* spp. and examine the role of (multiple) host association in genotypic clustering, patterns of migration and gene flow, by comparing the genetic diversity of isolates recovered from the UK and NZ. Bacterial host interactions are further investigated in chapter seven using an insect model of virulence, by evaluating potential differences in virulence gene profile between carriage and asymptomatic *S. aureus* isolates. The association of virulence with *agr*, the global gene regulator, and the relevance of this association is also examined. Finally chapter eight briefly reviews the major findings of this thesis and attempts to draw conclusions to bacterial host interactions.

*"Perseverance is the hard work you do after you get tired of doing the hard work you already did."*

Newt Gingrich
Speaker of the U.S. House of Representatives

*"Support bacteria! It's the only culture some people have."*

American bumper sticker

## ACKNOWLEDGEMENTS

First of all, I am indebted to my supervisor Ed Feil for the guidance, patience, friendship and support he has provided, throughout my ups and downs, during the course of this work. Similarly to Jess Cooper, an amazing friend and colleague, for her support and encouragement over the years. Thanks also to Sandra Barns for her invaluable assistance with the *Manduca*, teaching me everything there is to know about caterpillars, from life-cycles to mood swings, and to Paul Wilkinson for his sequencing expertise.

Many of my colleagues and contemporaries have helped me along the way. Thanks also to the many cider (& normal beverage) drinking buddies I have had over last few years; Andrea Dowling, the provider of cheeky pints and moral support; Sam Boundy for his continual pestering of "Are you nearly finished yet?"; Pete Millichap for IT support and Yorkie bars; Fiona McDonald and John Steventon for long-distance support. Amelia, Caroline, Nic, Yannis, Ronald, Rob and Michelle have been the source of ideas and amusement, as has Kevin with his invaluable IT assistance and patience reminding me how to use VPN. You are all my family away from home. I would especially like to thank Mr Mirghani and Mr Nassar at Monklands Hospital for stapling me back together again, to live on and fight another day.

Finally, I would like to thank those closest to me, whose presence helped make the completion of my PhD possible. Thanks to Reto for his support, love and encouragement. Most importantly, I would like to thank my parents and sister whose love, continual support and encouragement have helped me enormously over the last few years, always knowing when I need them the most; it is to them I dedicate this thesis.

## CONTENTS

# CHAPTER ONE

# MICROBIAL DIVERSITY & GENOMICS

# 1. INTRODUCTION

## 1.1 Biodiversity & population biology

### - What is microbial biodiversity?

Biodiversity is the variety and abundence of organisms which exist on Earth: the different plants, animals and microorganisms, their genes and the ecosystems of which they are a part. The classification of Earth's biodiversity has been a labour of love throughout the $18^{th}$ and $19^{th}$ centuries, with biologists having long toyed with the unifying idea of a common ancestor from which all life has evolved. The first known prokaryotic microfossils are dated to 3500 MYA (million years ago) (Kutschera and Niklas, 2004). Microbial organisms are pervasive, ubiquitous, and essential components of all ecosystems. The geochemical composition of Earth's biosphere has been moulded largely by microbial activities (Fig. 1). Yet, despite the predominance of microbes during the course of life's history, only a small representation, < 1 %, of the true microbial diversity (approximately $10^8$ bacterial species) have been identified as the vast majority of species are currently unculturable (Sherratt, 2001). The total number of prokaryotic cells has been estimated to be approximately 4-6 × $10^{30}$ (Whitman *et al.*, 1998).

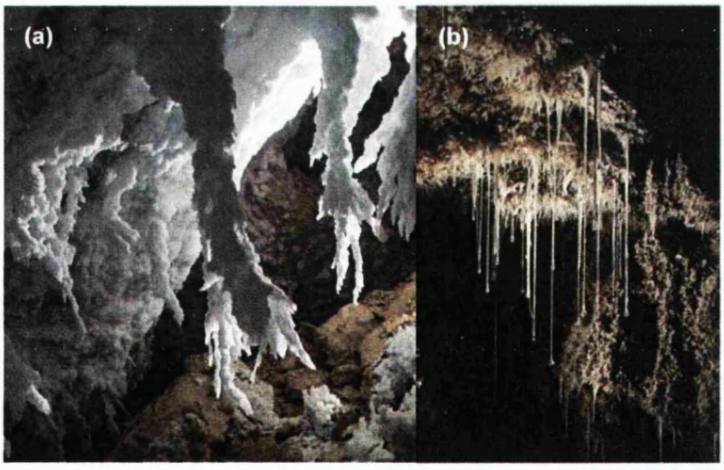



**Figure 1** Lechuguilla Cave in New Mexico (a) has all the hallmarks of being carved out by sulphuric acid, a by-product of bacterial metabolism. In 1998, scientists discovered sulphur-oxidizing bacteria (b) and the sulphuric acid they created in Mexico's Cueva de Villa Luz, still living on the cave ceiling and walls forming structures called snottites (Attenborough, 2006; Northup *et al.*, 2003).

The description of the composition and structure of microbial populations and communities is an important starting point in studies of microbial biodiversity, and this field has grown markedly since the creation of the DIVERSITAS International Research Programme in 1996 (DIVERSITAS, 2006). Microbial biodiversity has received particular attention in areas where industrial applications are evident, such as for marine, medical, and food biotechnology, and where microbial activity has important implications for Earth's climate and for the bioremediation of polluted sites (Hilger and Humer, 2003; Morris *et al.*, 2002). Much can be learned about bacterial evolution by studying isolates recovered from the wild or by evolving bacterial experimentally.

-Relevance of bacteria to the study of population biology

Population biology is the study of biological populations of organisms, investigating their diversity, dynamics, ecology and genetics. A population, can be defined at various spatial scales; local, regional and global. Local populations can occupy very small habitat patches, for example, one wood, meadow or lake (Mader, 1996). Regional populations depend on colonisation from adjacent areas to form metapopulations. Global populations can be considered at a scale of regions, islands, continents or seas to include the activity of one population (Storch, 2001). Population genetics is considered the most important, fundamental body of theory in evolutionary biology and is the study of the allele frequency, distribution and change under the influence of evolutionary forces: natural selection, mutation, and gene flow (these will be discussed later in more detail) (Mader, 1996; Page and Holmes, 2000). It also takes account of population subdivision and population structure, and attempts to describe both quantitively and qualitatively the variation that exists within natural populations, and to understand the processes involved in the generation of variation, i.e., recombination, mutation and lateral gene transfer (Page and Holmes, 2000).

Prokaryotes are excellent model organisms for studying population biology for a number of reasons. For example, they are adaptable and can evolve very quickly to changing environmental conditions; have large population sizes ensuring the absolute amount of mutational variation entering the population is at its greatest (Jordan *et al.*, 2002); a short generation time makes it possible to observe population dynamics over thousands of generations, for example, *Escherichia coli* grown on glucose salts has a

generation time of 17 minutes; bacteria have a haploid genome, therefore they possess only one copy of each gene and any mutations are immedialety subjected to natural selection; they have the capacity for gene exchange within and between populations by lateral gene transfer; the relative ease of gene manipulation; and the maintainence of bacterial stock cultures.

## 1.2 Factors which drive microbial diversity

### -Bacterial genomics

In prokaryotes, the genome typically consists of a single chromosome of double-stranded DNA, which is typically circularized although a few species (e.g. *Borrelia burgdorferi*) have a linear chromosome. Some species have two circular chromosomes per cell, e.g. *Vibrio cholerae* (Schoolnik and Yildiz, 2000). Bacterial genome size ranges from 0.16 megabases (Mb) to 9.1 Mb and the G+C content can vary from 25% to 75%. In most cases, bacteria belonging to the same species and/or closely related species have very similar G+C content and generally are of similar size (Chan, 2003). However, for some bacteria of the same species, the size of the genome can vary markedly, for example, *E. coli* K-12 has a genome size of 4.6 Mb, whereas *E. coli* O157:H7 is 5.6 Mb (Chan, 2003). Bacterial genomes typically contain 2000-4000 genes, but only a small fraction (~15%) of these genes are essential for the viability of the cell under optimum laboratory conditions. The close correlation between genome size and the number of genes suggests that bacteria do not contain large amounts of non-coding DNA. In the vast majority of complete bacterial genome sequences, protein coding genes account for 87 – 94 % of the genome.

The bacterial genome is broadly composed of two types of genes; core "essential" genes and "accessory" genes (Fig. 2). The core "essential" genes are very common or ubiquitous to all eubacteria and are indispensable for the survival of the cell. These include genes which encode for DNA replication, transcription and translation, cell envelope structure and key metabolic pathways. The core genes are considered to relatively stable over evolutionary time and are unlikely to be subject to positive selection for variation or any kind of strong directional selection. Therefore these genes are suitable markers for determining the degree of bacterial evolution (Feil and

Spratt, 2001; Santos and Ochman, 2004). This does not however mean that the core genes are immune to horizontal gene transfer (HGT) as low level recombination is observed frequently in genomic studies (see Chapters 3, 4, 5, and 6 for evidence of recombination within core "essential" genes). Accessory genes, may or may not be present in a given strain and typically fall into one of three classes; i) genes that encode secondary metabolic pathways, ii) genes such as transposases, that are associated with mobile elements (e.g. phage, plasmids), and iii) genes encoding products that are involved in interactions with the external environment (Feil, 2004). They often encode for bacterial pathogenicity and antibiotic resistance. Accessory genes are dispensable genes, that is, they are not necessarily essential for the survival of the bacteria, and can be deleted from the genome without an appreciable loss of fitness. As they are expendable, accessory genes are frequently not native to the genome and have been introduced by horizontal gene transfer (Feil, 2004).

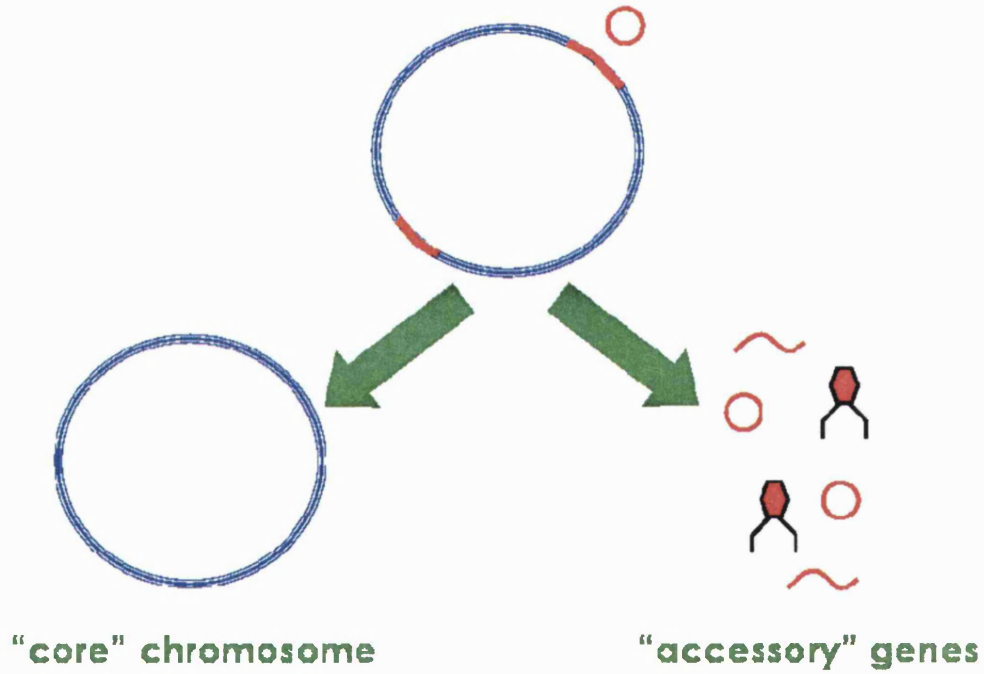



**Figure 2** Composition of bacterial genomes. Bacterial genomes can be considered as two parts, i) core chromosomal genes and ii) accessory genes which have been introduced to the genome by horizontal gene transfer.

-Genome replication

Genome replication occurs in three phases; initiation, elongation and termination. Among bacteria, DNA replication is best understood in *E. coli* (Messer, 2002). The initiation of replication begins at the origin region (*ori*), a region of DNA believed to be associated with the cell membrane (Brooks *et al.*, 1998; Turner *et al.*, 2001). Initiator proteins (DnaA) form a complex of 30 – 40 molecules around which *ori* DNA becomes wrapped, which facilitates the loading of DNA helicase, single-stranded binding protein (Ssb) and DNA primase required to form replication forks (Turner *et al.*, 2001). As part of the parental DNA unwinding process, positive supercoiling occurs, however this is relieved by the topoisomerase (type II) DNA gyrase (Messer, 2002).

During bi-directional replication both strands of the double helix must be copied. This is an important complication because DNA polymerase enzymes are only able to synthesize DNA in the 5'→3' direction. This means that one strand of the parent double helix, the leading strand, can be copied in a continuous manner, however replication of the lagging strand has to be carried out in a discontinuous manner. This results in a series of short segments (called Okazaki fragments) that must be ligated together to produce the intact daughter strand. A second complication within the replication process arises because template-dependent DNA polymerases cannot initiate DNA synthesis on a molecule that is entirely single-stranded: there must be a short double-stranded region to provide a 3' end onto which the enzyme can add new nucleotides (Brown, 2002). RNA primers are synthesised, one to initiate complementary strand synthesis on the leading polynucleotide, and one for every segment of discontinuous DNA synthesized on the lagging strand. Both leading and lagging strand primers are elongated by DNA polymerase III holoenzyme, a multi-subunit dimer complex, one half of which acts as a polymerase, and the other half a 3'→5' proofreading exonuclease (Brown, 2002; Turner *et al.*, 2001). DNA polymerase III is released from the lagging strand once the primers have been elongated, and is replaced DNA polymerase I which has 5'→3' polymerase and 5'→3' exonuclease activity. The exonuclease activity of DNA polymerase I removes the primers while the polymerase function simultaneously fills the gaps with DNA by elongating the 3' end of adjacent Okazaki fragments. DNA ligase joins the lagging strand fragments together (Turner *et al.*, 2001).

Replication of chromosomal DNA proceeds until the replication forks reach the termination site, *terC*, in the case of circular chromosomes, or chromosomal ends, in the case of linear chromosomes. Once replication is completed, the two daughter cells remain interlinked and must be unlinked by topoisomerase IV. They can then segregate into two separate daughter cells by movement apart from their membrane attachment sites.

-Mechanisms of variation

In eukaryotes genetic recombination occurs as a result of sexual reproduction, however bacteria reproduce asexually by means of binary fission. Therefore bacteria must utilise other mechanisms to introduce genetic variation into the population. Each of these methods will be discussed in turn:

i) Mutation

Although DNA replication is a very efficient and accurate system, it does not function correctly on every occasion; errors, or mutations can arise in the process. Point mutations are the most common type of mutation of which there are four basic types. They are substitution of a nucleotide for another nucleotide, deletion of nucleotides, insertion of nucleotides, and inversion of nucleotides (Dale, 1995). Nucleotide substitutions can be divided into two classes; transitions and transversions (Page and Holmes, 2000). A transition is the substitution of a purine (adenine or guanine) for another purine or the substitution of a pyrimidine (thymine or cytosine) for another pyrimidine (Fig. 3). Transversions occur when a pyrimidine is substituted for a purine, or vice versa. In most DNA segments, transitional nucleotide substitutions are known to occur more frequently than transversions (Brown, 2002; Gojobori *et al.*, 1982). Deletion, insertion and inversion may occur with one or more nucleotides as a unit. If this occurs in a protein coding gene, they may shift the reading frame of the nucleotide sequence so that all nucleotides downstream from the mutation will be improperly grouped (Page and Holmes, 2000).

In the case of protein coding genes, nucleotide substitutions that result in synonymous codons are called synonymous or silent substitutions. The mutation has no effect on the coding function of the genome; the mutated gene codes for

exactly the same protein as the unmutated gene. Changes that result in nonsynonymous codons are called nonsynonymous or amino acid substitutions, where the mutation alters the codon so that it specifies a different amino acid (Nei and Kumar, 2000; Page and Holmes, 2000). In addition, mutations may convert a codon that specifies an amino acid into a stop codon. This is a nonsense mutation and it results in a shortened protein because translation of the mRNA stops prematurely at the new stop codon rather than proceeding to the correct stop codon further downstream. If a mutation converts a stop codon into one specifying an amino acid, a readthrough can occur where the protein is extended by an additional series of amino acids at its C terminus (Brooks *et al.*, 1998). Because of the properties of the genetic code, most synonymous substitutions occur at the third nucleotide position of codons, but some can occur at the first position. All nucleotide substitutions at the second position are either nonsynonymous or nonsense mutations.

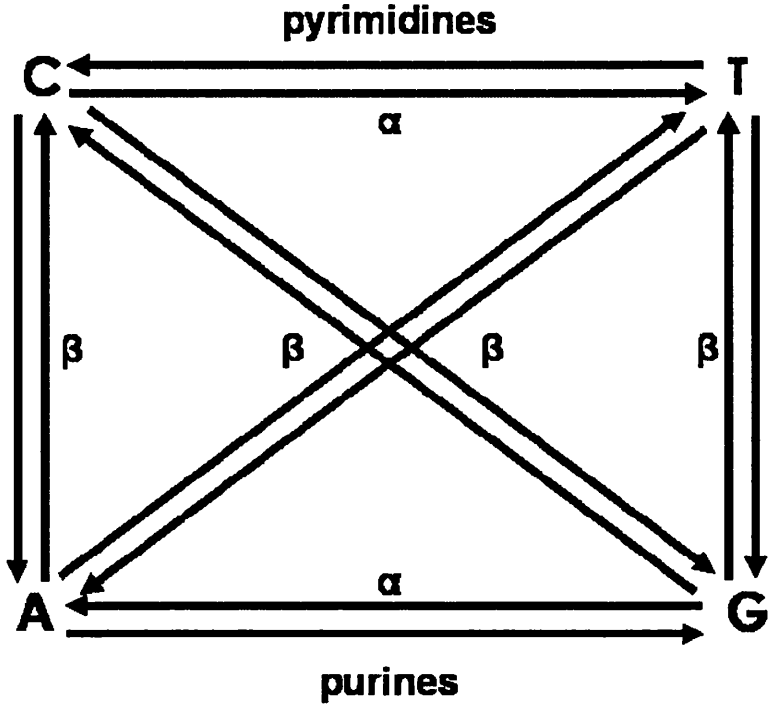



**Figure 3** Transitional (A←→G) and T ←→C) and transversional (A←→C) and (G←→T) nucleotide substitutions. α and β are the rates of transitional and transversional substitutions, respectively. (Nei and Kumar, 2000).

ii) Horizontal gene transfer

The natural transfer of genetic material between bacteria in the environment generates much of the genetic diversity required for natural selection and evolution. The introduction of novel genes or alleles by horizontal gene transfer (HGT) allows for niche-specific adaptation, which eventually might lead to bacterial diversification and speciation (Eisen, 2000; Jain *et al.*, 2002; Ochman *et al.*, 2000). Traditionally it was thought that microorganisms evolved strictly clonally, with vertical inheritance of DNA and little exchange of DNA amongst diverse species (Jain *et al.*, 2002). It has since been realised that HGT plays an important role in shaping bacterial genomes. Furthermore, the transfer of novel DNA into bacterial genomes is thought to be a major source of new phenotypic characteristics which allow survival under strong selection regimes (Jermyn and Boyd, 2005). It may even be that archaeal, bacterial and ancestral eukaryotic genomes have throughout time shared genomic information through HGT (Doolittle, 1999). Examples of HGT include transfers from organellar to nuclear genomes in eukaryotes, transfer of plasmids between bacterial species and the spread of genes conferring antibiotic resistance and virulence determinants (Eisen, 2000).

There are three principle mechanisms of recombination that facilitate the transfer of genes in prokaryotes (Jain *et al.*, 2002):

- conjugation, which involves the direct transfer of DNA from one cell to another;
- transformation, in which a cell takes up isolated DNA molecules from the environment surrounding it;
- transduction, in which the transfer of DNA is mediated by bacteriophages.

Not all bacterial species exhibit all of these modes of genetic transfer. Conjugation is most easily demonstrated in Gram-negative bacteria, such as *E. coli*, but does occur in some Gram-positive genera such as *Streptococcus* and *Streptomyces* (Dale, 1995). Transformation does not occur widely in the environment due to the presence of nucleases which degenerate extracellular DNA. Some species of bacteria are naturally transformable (*Neisseria gonorrhoeae, Streptococcus*

*pneumoniae* and *Haemophilus)* and present convenient models for biotechnology applications (Provvedi *et al.*, 2005), however in many other species transformation is only readily demonstrated after manipulation of the cells (e.g. chemical treatment). Horizontal gene transfer provides a mechanism by which bacteria can constantly vary their surface antigens to avoid recognition by host defences.

Conjugation is the transfer of DNA from a donor to a recipient by direct physical contact between the cells. The mechanisms of transfer vary from one case to another, but are best understood in *E. coli*. Two mating types occur in bacteria, i) donor (male) and ii) a recipient (female). To be a donor, a bacterium must have a F-plasmid designated $F^+$, which confers certain donor characteristics upon cells; a sex pilus, an extracellular protein extrusion that attaches donor cells to recipient cells lacking fertility factors (Dale, 1995). A cytoplasmic bridge between the cells allows a single strand of $F^+$ plasmid, synthesised by the donor, to pass to the recipient cell ($F^-$), where the complementary strand of DNA is formed (Prescott *et al.*, 1996). Transfer replication is via a special mechanism using rolling circle replication from *oriT,* not in synchrony with the bacterial chromosome. Occasionally, the plasmid integrates into a random position in the bacterial chromosome creating a Hfr (high-frequency recombination) bacterium. Hfr bacteria are still able to initiate conjugation with $F^-$ cells, but the outcome is different from conjugation involving $F^+$ bacteria. DNA transfer is initiated at *ori-T* on the integrated plasmid and the Hfr cell tries to transfer the entire bacterial chromosome, plus plasmid to the recipient cell. Usually only partial chromosome transfer occurs, with the recipient receiving only part of the plasmid and remains $F^-$ (Brooks *et al.*, 1998; Prescott *et al.*, 1996). Any chromosomal DNA received by the recipient cell can undergo recombination because the DNA sequences will be homologous.

Transformation involves the uptake of exogenous DNA and its incorporation into the recipient chromosome in a heritable form. Approximately 90 transformable bacterial species have been identified (e.g. *Streptococcus*, *Bacillus*, *Haemophilus*, *Neisseria* and *Pseudomonas*), but not all are known to be competent for DNA uptake in the natural environment (Prescott *et al.*, 1996; Sorensen *et al.*, 2005). In natural transformation the DNA comes from a donor bacterium. This process is random, in the sense that any portion of a genome may be transferred between bacteria. When

bacteria lyse, DNA is released into the surrounding environment. The DNA fragments may be relatively large and contain several genes. If a fragment comes into contact with a competent cell, one able to take up DNA and be transformed, it can be bound to the cell and taken inside (Sorensen *et al.*, 2005). The state of competency is growth-phase dependent, for example, *S. pneumoniae* is only competent during its exponential phase at cell densities of $10^7$ to $10^8$ cells per ml (Prescott *et al.*, 1996). The mechanism for transformation varies between bacterial species and is best understood in *S. pneumoniae*. A competent cell binds a double-stranded DNA fragment with the aid of DNA binding proteins. One strand is completely degraded by envelope-associated exonucleases, whilst the other having been hydrolysed into small fragments, moves through the plasma membrane (Dale, 1995). The single-stranded fragment can then be stably maintained either by homologous recombination or by recircularisation of plasmid DNA (Sorensen *et al.*, 2005). In many organisms, including *E. coli*, competent cells do not occur naturally. Artificially inducing a competent state (e.g. treatment of cells with $CaCl_2$) can overcome this problem (Griffiths *et al.*, 2000).

Transduction is a specific HGT process as bacteriophages have a limited host range, sometimes being limited to a single bacterial species. However this does not restrict bacteriophages acting as gene-transfer vehicles, introducing DNA from donor bacteria into recipient bacteria by infection. Instead, their great abundance (Jiang and Paul, 1998) and their temperate ability to insert themselves into chromosomes as a prophage without causing cell lysis, make them very efficient at altering the genetic content of their hosts (Davison, 1999). Transduction may be either "generalized" (e.g., by coliphage P1), whereby any bacterial gene may be transferred, or "specialized" (e.g., by coliphage lambda), where only genes which flank the prophage are transferred (Jiang and Paul, 1998). In generalized transduction, bacterial DNA is accidentally incorporated into the phage capsid. The transducing particle is capable of infecting a recipient cell, since the information necessary for attachment and injection of DNA is carried by the phage and not by the nucleic acid it contains. As in transformation, the DNA must be incorporated into the recipient cell's chromosome by homologous recombination (Griffiths *et al.*, 2000). In specialised transduction, during excision of the prophage, an error occurs where some of the host DNA is excised with the phage DNA; genes which flank the

prophage. As the bacteriophage replicates, the segment of bacterial DNA replicates as part of the phage's genome. Every phage now carries that segment of bacterial DNA. After the release of phage and infection of a recipient, lysogenisation of the recipient bacterial cell can occur resulting in the stable transfer of donor genes. Along with the prophage, the bacterial DNA integrates with the recipient bacterial chromosome (Dale, 1995; Sorensen *et al.*, 2005).

The transfer of genetic material by the above means is thought to blur the boundaries of species and to have long-term evolutionary consequences because the acquisition of novel virulence genes or novel metabolic properties can change the niche of an organism (i.e. from soil dwelling to host) and can alter pathogenicity (Feil and Spratt, 2001). This will be discussed below in further detail. HGT brings new genes into the genome that are either homologous to existing genes or are entirely new sequence families (Ochman, 2001). The potential for acquiring and replacing an existing gene has been thought to generally decrease with the phylogenetic distance between the donor and recipient lineages, with an acquired gene or gene segment more likely to be beneficial if it has some properties similar to the recipient genome (Ochman, 2001). However, despite this decrease in recombination frequency with increasing phylogenetic distance, there are reports of occasional HGT between kingdoms, i.e. between archaea and eubacteria (Aravind *et al.*, 1998; Gophna *et al.*, 2004).

Since the realisation of the importance of HGT in bacterial evolution, there has been much debate concerning which types of genes are most frequently transferred. The complexity hypothesis, proposed by Jain *et al.*, (1999), put forward the idea that not all genes are equally likely to be transferred and that extensive HGT occurs for operational genes (housekeeping genes), whereas informational genes (transcription and translation genes) are less likely to be involved with HGT. Because informational genes are part of complex essential systems, they are less likely to be transferred than operational genes (Jain *et al.*, 1999).

iii) Host adaptation

Selective pressures in host adaptation are the result of a continuous conflict between the divergent interests of each partner (Combes and Theron, 2000). The fitness of the host is increased by avoidance of the parasite (in this case the bacterium) and/or

its destruction, whereas the fitness of the parasite (reproductive success) is increased by a higher frequency of encounters with susceptible hosts and enhanced survival rate following infection. These selective processes are in agreement with the Red Queen hypothesis (Ochoa and Jaffe, 1999), which assumes indefinite adaptive changes in both partners in order to counter-act measures against the other (Combes, 2000). A number of factors, including genetic (e.g. HGT), environmental (e.g. nutritional status) and external (e.g. prevention, vaccination, therapy) help determine the outcome of this evolutionary arms race (Woolhouse *et al.*, 2001). Understanding microbial strategies for adaptation is the key to understanding pathogenesis.

It is widely assumed that bacterial pathogens involved in co-evolution are generalists (Combes and Theron, 2000; Woolhouse *et al.*, 2001), that is they have the ability to adapt to a wide range of environmental conditions. However, specialists can arise on each occasion a specific virulence factor is selected for against the host, resulting in a small number of predators possessing adaptations which allow them to rapidly specialise in the host (Combes and Theron, 2000). This can result in clonal expansion within the host and low level diversity within the species. Many important human, domestic and wildlife diseases are caused by pathogens which have the capacity to infect multiple host species; this includes the 60% of human pathogen species that are zoonotic (Shrivastava *et al.*, 2005; Woolhouse *et al.*, 2001). Not only can they infect multiple species, but pathogens such as influenza A (Baigent and McCauley, 2003), *Burkholderia* spp. (Valvano *et al.*, 2005) and *Staphylococcus aureus* (Smyth *et al.*, 2005) can infect hosts not only from different species, but from different Orders or Classes. Furthermore, host switching has contributed to the rapid emergence of new strains, for example, Simian Immunodeficiency Virus (SIV), which is thought to have been transmitted to humans from other primates, has diverged rapidly into new single-host pathogens, HIV-1 and HIV-2 (Hahn *et al.*, 2000)

Genetic variability of pathogens is of great importance in shaping host adaptation. This may involve a small number of point mutations, e.g. Wβ and γ phages (Schuch and Fischetti, 2006), or more major genetic changes, such as genome rearrangements, e.g. *Burkholderia* spp. & *Vibrio* spp. (Coenye *et al.*, 2005) or the acquisition of genetic elements (sometimes associated with virulence as well as host

specificity), e.g. meticillin-resistant *Staphylococcus aureus* (Kuroda *et al.*, 2001). Genomic flux, the acquisition of new genes by gene duplication or HGT, and the deletion of non-essential genes, plays an important role in host adaptation. *Buchnera*, the endosymbiont of aphids, has undergone extensive genome reduction to become a highly specialised intracellular bacterium that is totally dependent on its host's protective environment for survival, as is its host for essential nutrients which is cannot synthesize (Shigenobu *et al.*, 2000; Wren, 2000).

The environment is in constant flux, often fluctuating between beneficial and detrimental conditions for organisms. Bacteria can respond quickly to environmental challenges by for example, increasing mutation rates (Horst *et al.*, 1999). Low levels of mutator cells in a population can act as a reservoir of variation to allow for selection of the fittest in times of stress. Such cells can mutate rapidly to accumulate the multiple mutations required for selection and the mutator population can expand rapidly for a particular trait, or by changing the expression of genes which play a specific role in the interaction between the bacterium and the host. The genome sequencing of *Bifidobacterium longum* revealed an excessive number of genes associated with oligosaccharide metabolism. These genes were amplified by both gene duplication and HGT, as a response to strong environmental pressure to amplify and diversity of *B. longum*'s metabolic capabilities, in response to competition for varied substrates in the human gastrointestinal system (Schell *et al.*, 2002). Intracellular bacteria that establish lifelong interactions with their hosts often increase the number of genes involved in the generation of antigenic variability. Functional pseudogenes in the *msp*2 and *msp*3 superfamilies of *Anaplasma marginale* result in constant genetic and antigenic variants, allowing lifelong infections of cattle (Brayton *et al.*, 2005; Sallstrom and Andersson, 2005).

-Gene expansion

There are three mechanisms by which genomes expand their repertoire of genes; HGT, gene duplication and gene genesis. Gene duplication has been thought of as amplifying the number of copies of genes so that the copy/copies are free to evolve and diverge until it attains a novel function (Hooper and Berg, 2003). Hooper and Berg (2003) suggest that new paralogs, i.e. new gene copies, avoid neutrality by gene

amplification effects. In addition, not only can entire genes be duplicated, but also gene segments and entire genomes (Hooper and Berg, 2003; Wolfe and Shields, 1997). Recently duplicated genes often have relaxed evolutionary constraints, which allow the duplicated genes to functionally diversify and promote biochemical innovation through mutation and recombination (Wagner, 2001). The fate of the duplicated gene is i) positive selection if the duplication is beneficial or the mutations it has developed now enable the organism to perform a novel task, ii) redundancy of the new paralog, i.e. the gene becomes a pseudogene or is deleted or iii) both paralogs are selectively maintained by subfunctionalisation (Hooper and Berg, 2003; Zhang and Kishino, 2004). Gene genesis is the *de novo* origin of a gene. It can be defined as occurring in the lineage leading to the most recent common ancestor of the species in which the orthologous genes are present (Snel *et al.*, 2002).

-Genome reduction

Currently, 324 bacterial genomes are available, of which 36 % have a genome smaller than 2 Mb (Hallin and Ussery, 2004). Initially it was believed that the smallest of these genomes represented ancestral genomes, however it is now clear that these genes are derived from larger genomes through gene loss events (Moran and Mira, 2001). Selective pressures can lead to genome reduction in order to promote efficiency or competitiveness during replication (Mira *et al.*, 2001). Therefore gene loss by large deletion events or by frameshift mutation serves to counterbalance an increase in gene number from HGT and gene duplication events, removing harmful mutations and provide a selective advantage. An excellent example of bacteria which have conferred a selective advantage from gene loss is that of *Shigella* spp. When this genus evolved from *E. coli* to become pathogenic, it not only acquired virulence genes by HGT, but also lost genes via the deletion of a large genomic segment. The gene deletions inferred selective pressures that were detrimental to a pathogenic lifestyle, therefore the loss provided an evolutionary pathway that enabled enhanced virulence (Maurelli *et al.*, 1998). Extreme genome reduction, producing genomes of 1.5 Mb or less, has been documented in several bacterial groups with a host-associated lifestyle (obligate parasites and symbionts), including *Buchnera*, *Chlamydiae*, *Mycoplasma* and *Wolbachia* (Cohan, 2002; Mira *et al.*, 2001). One clear basis for genome reduction is that bacteria living

continuously in hosts can obtain many compounds of intermediate metabolism from the host cytoplasm or tissue, thereby being able to discard their corresponding biosynthetic pathway and genes (Moran, 2002). For instance, the obligate intracellular pathogen *Chlamydia trachomatis* lacks many biosynthesis capabilities, but has retained loci for the interconversion of metabolites obtained from host cells (Stephens *et al.*, 1998; Wren, 2000). An exception to this rule is with the obligate symbiont *Buchnera* which compared to other symbionts, has a substantial increase in the number of loci which specify the biosynthesis of essential amino acids needed by its host, but has lost pathways for amino acids that the host can produce itself (Shigenobu *et al.*, 2000).

There are two possible mechanisms to explain genome reduction. The first is a continuous reduction process, with genes being lost individually through a large number of small deletion events (Silva *et al.*, 2001). The second is that many genes were lost simultaneously at an early stage of the internalisation process through the fixation of a few deletion events that spanned multiple genes and were associated with many rearrangement events (Moran and Mira, 2001; Tamas *et al.*, 2002). The latter hypothesis is the most commonly preferred.

-Chromosomal rearrangements

Genome rearrangements are another driving force behind bacterial genome evolution. As more bacterial genomes have been sequenced, it has become apparent that the level of conservation can be high when organisms are phylogenetically closely related, but that gene order conservation is lost as phylogenetic distance increases (Coenye *et al.*, 2005). However this is not the case for all microbes. Comparison of the *Wolbachia* *w*Mel and *w*Bm genomes revealed extensive genome shuffling had taken place, most likely due to the large number of DNA repeat and mobile elements present in these genomes (Foster *et al.*, 2005; Wu *et al.*, 2004). In contrast to other anciently host-restricted lineages, multiple *Wolbachia* infections of arthropod hosts can occur, where they undergo recombination and potentially spread elements between different strains (Jiggins, 2002; Mouton *et al.*, 2004; Werren and Bartos, 2001). As a result of the high levels of mobile elements, it may be possible that *Wolbachia* has many lineage-specific chromosomal rearrangements (Wu *et al.*,

2004). The degree of genome flexibility in bacteria is dependent on the content of repeated and mobile DNA sequences such as insertion sequence elements, transposons, plasmids and bacteriophages. Chromosomes containing a higher repeat density have higher rates of rearrangements leading to accelerated loss of gene order (Coenye *et al.*, 2005).

## 1.3 Adaptive evolution

Adaptive evolution is a process whereby mutations at the genetic level that are beneficial to an individual or population, are retained and fixed by positive selection (Hall, 2004; Sharp, 1997). In some cases, the mutation may be neutral, and pure chance determines whether the mutation will be driven to fixation. Neutral mutations may build up to a high frequency in a population and may represent a hidden source of variation important to the population when the environment changes (Hall, 2004). If a mutation is deleterious to the organism then it may be removed by purifying (negative) selection, thus ensuring that functionally important genes rarely change. If altered environmental conditions occur, then some mutations may confer an advantage, giving the organism a greater chance to survive. Under these circumstances, selective pressures favour the mutation to be retained and it may become fixed in the population (Hastings and Rosenberg, 2002). This phenomenon is called diversifying (positive) selection (Hall, 2004). Adaptive evolution is thought to be a transient event in protein-coding genes that frequently leads to the generation of novel functions; e.g. alteration of surface proteins on pathogenic bacteria or viruses to avoid detection from the host immune system (Brayton *et al.*, 2005; Smith *et al.*, 1995).

Adaptive evolution in protein coding genes can be detected from the comparison of synonymous and nonsynonymous substitution rates (Nei, 2005). Synonymous substitutions are usually regarded as neutral, or at least having a much smaller effect on fitness than nonsynonymous substitutions. Adaptive evolution is thought to be the result of selective pressures favouring a high level of nonsynonymous substitutions (Nei, 2005).

## 1.4 Variation in bacterial genomes

Microbial diversity results from shifting levels of clonality and recombination, within and between species; each is as important as the other in shaping bacterial populations. Bacteria such as *Salmonella enterica* and *Staphylococcus aureus* are considered clonal (Feil *et al.*, 2003; Smith *et al.*, 1993; Spratt *et al.*, 2004), whereas other bacteria, such as *Helicobacter pylori* and *Neisseria meningitidis*, are non-clonal (Feil and Spratt, 2001). Clonal populations are a consequence of vertical transmission, and therefore the absence of genetic exchange (Smith *et al.*, 1993). In a clonal population, each member is genetically identical to its mother cell and genetic changes that occasionally arise are limited to the descendents of the bacteria in which they occur, resulting in a distinct genetic lineage (Gupta and Maiden, 2001). These changes can arise by either point mutation or recombination, but both have the same outcome, that is, a change from the ancestral genotype. If the recombination:mutation ratio is low, then the population will be clonal. The higher the ratio becomes, the more non-clonal the population becomes. Clonal populations can be accurately modelled by congruence between tree-like phylogenies, and exhibit high levels of linkage disequilibrium; a non-random association among alleles at different loci (Feil *et al.*, 2003).

Non-clonal populations are generated by frequent recombination events are typically characterised by linkage equilibrium; a random association among alleles at different loci (Feil *et al.*, 2003). Non-clonal populations can be accurately modelled by incongruence between tree-like phylogenies and are better represented by network-like phylogenies such as splits decomposition (Bandelt and Dress, 1992; Huson and Bryant, 2006). Homologous and non-homologous recombination can have striking effects on microbial sequence diversity, with genomes from different strains of the same bacterial species showing marked differences in gene content within a single species. Some genomic differences within a species may be the result of gene loss since the emergence of the species, however many are due to gene expansion by non-homologous recombination and can confer a selective advantage (Spratt *et al.*, 2001). For example, non-homologous recombination has resulted in the acquisition of a number of virulence factors which have been responsible for the emergence of a pathogenic strain of *E. coli* (Reid *et al.*, 2000).

Frequently, clonal expansions take place on a freely recombining "background" population (Smith *et al.*, 2000). This process leads to a small number of genotypes dominating the population (Fig. 4). However, these dominant clones might subsequently diversify by mutation or recombination (Gupta and Maiden, 2001) and lose genetic cohesion.

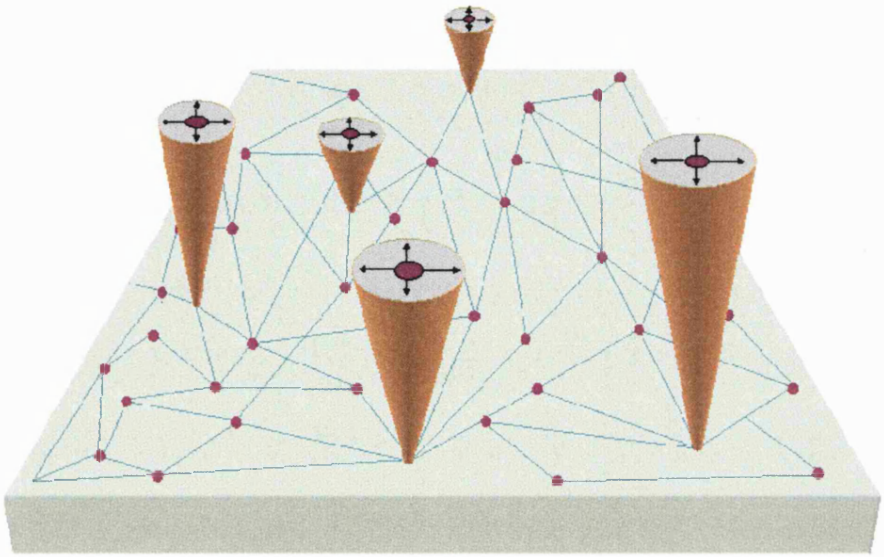


**Figure 4** The delicate relationship of recombination and clonality. Such a population is composed of two parts, i) the recombining "background", which is composed of a large number of relatively rare and unrelated genotypes (small pink circles) that are recombining at a high frequency, and ii) clones which emerge from a single highly adaptive genotype (purple circles) that increase in frequency under selection to reach an observable frequency within the population. The diversification of these clones (indicated by the arrows) is predominantly driven by recombination but also by mutation (Smith *et al.*, 2000).

## 1.5 Biological species concept

Species are one of the fundamental units of comparison in virtually all fields of biology. The species concept for asexual bacteria is fundamentally different from the species concept for eukaryotes (Mader, 1996). The most popular eukaryotic

species concept, the Biological Species Concept (BSC), pioneered by Ernst Mayr, is as follows:

*"...a species is an array of populations which are actually or potentially interbreeding, and which are reproductively isolated from other such arrays under natural conditions."*

(Mayr, 1942).

Mayr has revised this throughout the years, however reproductive isolation is still the centrepiece of the BSC. The definition of a species might seem clear cut for eukaryotes, barriers to gene exchange provide both a cause and a definition of species, however there are limitations to the concept for bacteria, which reproduce asexually. Considered outside a broad population and ecological context, such definitions are not biologically meaningful. All too often new bacterial species are assigned on the basis of a single isolate, which can further cause confusion (Cohan, 2002; Gevers *et al.*, 2005; Lan and Reeves, 2001).

Traditionally, microorganisms have described and classified according to their phenotype. Unfortunately, such characters provide little information about the evolutionary relatedness, which ideally should be the basis of any classification system (Hugenholtz and Pace, 1996). The majority of human pathogens have been named according to the disease they cause, for example, *Vibrio cholerae* is the pathogenic bacterium which causes cholera and *Neisseria meningitidis* causes bacterial meningitis.

The emergence of adaptive clusters and their relevance for systematics has recently been discussed by Gevers *et al.*, (2005). Clones corresponding to specific ecological adaptations (ecotypes) potentially provide a means for a universal species concept that can be applicable to both prokaryotes and eukaryotes (Cohan, 2002; Gevers *et al.*, 2005). Ecotypes are defined to be free to diverge without the constraint of one another's periodic selection event. They will typically be highly specialised, such that names "species" will correspond to a large number of ecotypes (Fig. 5). Furthermore, ecotype may not always correspond to single species clusters, as

cohesion may be confounded by large-scale migration (bringing together of previously separate clusters) or recombination (Cohan, 2002).

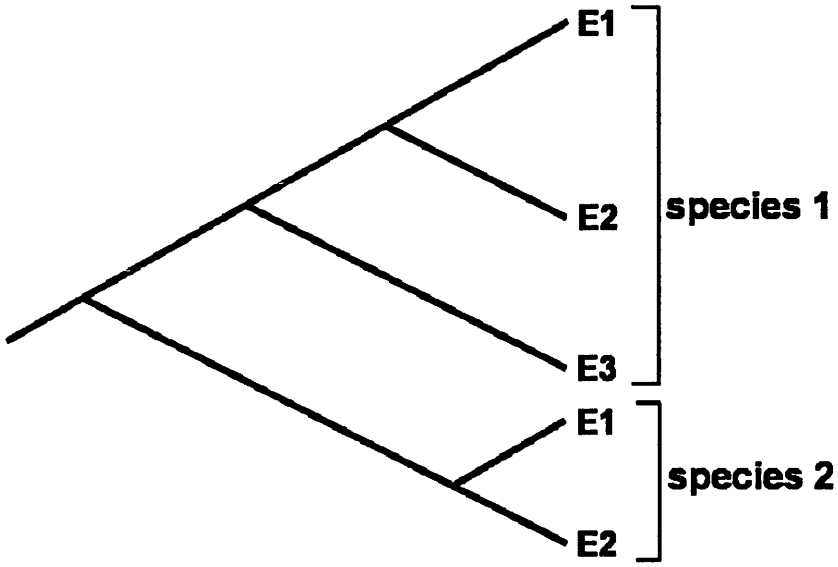



**Figure 5** Multiple ecotypes can be present within a single species. Ecotypes are labelled as E.

There are a number of traditional methods for bacterial taxonomy. As described previously, pathogenic bacteria were first classified according to the symptoms and disease they produced. Later, biochemical tests and serological properties were used but these tests do not provide adequate resolution between some named species, e.g. *Shigella* and *E. coli* (Lan and Reeves, 2001). Polyphasic taxonomy compiles different kinds of data and information into a classification of biological entities that contains a minimum of contradictions (Vandamme *et al.*, 1996). This includes phenotypic data (e.g. biochemical tests), genotypic data (e.g. DNA fingerprint data) and phylogenetic information (e.g. rRNA sequences) (Gevers *et al.*, 2005). At present, a bacterial species is:

*"...a category that circumscribes a (preferably) genomically coherent group of individual isolates /strains sharing a high degree of similarity in (many) independent features, comparatively tested under highly standardised conditions."*

(Stackebrandt *et al.*, 2002).

Since the 1970's the cornerstone of genotypic characterisation has been the measurement of overall genetic similarity among isolates, assessed by the degree to which their genomes hybridize under standard conditions (DNA-DNA hybridization). Isolates sharing greater than 70 % DNA-DNA relatedness, with less than 5 % difference in their melting temperature (ΔTm) are considered the same species, whereas isolates that share less than 50 % DNA-DNA relatedness belong to a different species (Gevers *et al.*, 2005; Vandamme *et al.*, 1996). DNA-DNA hybridization does however have limitations. Firstly, the procedure is very time consuming and is performed by only a select few laboratories, therefore is not suited for the rapid identification of prokaryotes. Secondly, the 70 % cut-off point for species demarcation in some circumstance is too low; some single "species" can differ by up to 20 % of their chromosome (Doolittle, 1999). This was the case with *Burkholderia mallei* and *Burkholderia pseudomonas*, which by DNA-DNA hybridization criteria were shown to represent a single genomic species, however through major differences in their biochemical activities, clinical symptoms and epidemiology represent two individual species (Rogul *et al.*, 1970).

Comparative analysis of the 16S rRNA gene is used extensively to determine the phylogenetic position of putative species. Developed by Woese (Woese, 1991), ribosomal RNA (rRNA) gene sequence similarity allows the classification of bacteria using a universally conserved trait; bacteria are considered a species where 16S rRNA gene sequence identity > 97% (Lan and Reeves, 2001). The technique allows rapid identification of organisms when compared to a sequence database (e.g. Genbank) and can be used with currently unculturable organisms. There are however several shortcomings in using this technique. Firstly, the 16S rRNA spans a very small proportion of the bacterial genome and has a slow rate of evolutionary change, therefore does not allow satisfactory phylogenetic resolution. As such, some bacterial isolates with 100% identical 16S rRNA sequences do not fulfil the biochemical requirements to place them together in a single species (Fox *et al.*, 1992). Secondly, rRNA is not a protein-encoding gene, therefore if insertions and deletions are present, this can introduce problems for sequence alignments (Cooper and Feil, 2004; Santos and Ochman, 2004). Thirdly, rRNA genes can be present in more than one copy, e.g. *Vibrio cholerae* strain N16961 has eight 16S rRNA genes

(Heidelberg *et al.*, 2000), which could result in the sequencing and alignment problems. Finally, 16S rRNA is by no way immune to recombination (Lan and Reeves, 2001); evidence of recombination, and potentially HGT, has been observed within this loci in *Photorhabdus* spp. (Richard ffrench-Constant, Edward J. Feil; unpublished data).

## 1.6 Molecular typing methods

The need for better intra-specific and inter-specific resolution in bacterial populations has never been greater. A wide variety of typing techniques are available, however most have been superseded by sequencing-based approaches, e.g. multi-locus sequence typing (MLST). A selection of these is discussed below:

### -Electrophoretic methods

Pulsed Field Gel Electrophoresis (PFGE) uses the pattern of restriction fragments on a gel to identify 'types'. Unlike regular restriction fragment length polymorphism testing PFGE uses enzymes which cut infrequently yielding fewer larger fragments, which simplifies the analysis considerably (Weller, 2000). PFGE is based on the observation that during continuous field electrophoresis, DNA above 30-50 kb migrates with the same mobility regardless of size; these fragments are too large to be separated. If the DNA is forced to change direction during electrophoresis, different sized fragments separate from each other, therefore it allows the separation of significantly larger pieces of DNA than conventional agarose gel electrophoresis (Turner *et al.*, 2001). PFGE has many advantages and disadvantages. An advantage it has over sequence-based approaches is that it is sensitive to large-scale genomic re-arrangements, therefore is suited to species with many genomic islands, insertion sequences or other mobile elements. It also can be readily performed in the absence of detailed information about the genome. The main disadvantage of PFGE is reproducibility, as each laboratory will follow their own individual protocol for running the gels. The inferred relationships between strains remains slightly subjective, and so detailed analysis of large datasets is difficult.

Multi-locus enzyme electrophoresis (MLEE) employs similar principles to that of PFGE, however indexes variation within multiple protein-coding housekeeping

genes on the basis of differing electrophoretic mobilities of the gene products (Selander *et al.*, 1986). The advantage to MLEE is that results can be directly tied to genetic loci and therefore are amenable to genetic analysis (Selander *et al.*, 1986). The mobility of the enzymes varies due different charge variants. This can range from small changes in morphology to surface amino acid variations. For most species there are only a few variants at each locus and so often it is necessary to test 20 or more loci in order to build an electrophoretic profile and electrophoretic type. MLEE can be very laborious as a result of testing so many loci. As with PFGE, reproducibility can be a problem between laboratories.

-Genus specific loci

Using individual protein-coding loci for identifying species has become increasingly popular in recent years. Bacterial phylogenies can be built by identifying a suitable protein-coding, housekeeping gene which is thought to be robust to recombination, for example, *recA* has been used for some years now to identify *Vibrio* species. Sequencing only one gene greatly reduces costs and allows straightforward phylogenetic analysis. However, there are a number of difficulties with this approach. Firstly, a locus would need to be selected that was balanced between an evolutionary conserved gene, which would not have enough discriminatory power, and a gene which has undergone frequent HGT, which would obscure any phylogenetic resolution. Secondly, no single gene should be used as the sole marker in identifying and distinguishing between bacterial species (Christensen *et al.*, 2001). As discussed previously, even the core "essential" genes are not immune to HGT, e.g. 16S rRNA, therefore multiple loci are necessary to buffer against the effects of recombination. As will be discussed in Chapters 5 and 6, although *recA* is evolutionary conserved it may still undergo HGT, leading to erroneous species assignments.

-Multi-locus sequence typing

Multi-locus sequence typing (MLST) is a method for the genotypic characterisation of bacteria at the infraspecific level. It is based on the principles of MLEE, but variation at housekeeping loci is indexed directly by DNA sequencing (Maiden *et al.*,

1998). Approximately 450-500 bp internal fragments of (normally) seven housekeeping genes are sequenced and allelic profiles created from their mismatches. Groups of isolates with identical allelic profiles define strains or clones and form the basis for a genotypic classification system. Strains of a species can be clearly defined by this profile, which is referred to as a sequence type (ST). For example, in table 1 strains 1, 2 and 3 belong to ST1; strain 4 belongs to ST2; strains 5 and 7 belong to ST3; strain 6 belongs to ST4; strains 8, 9 and 10 belong to ST5. Clonal complexes (CC) are typically composed of a single predominant genotype with a number of closely related genotypes. By focusing on allelic profiles and STs rather than nucleotide differences, genetic events such as point mutation and recombination are weighted equally. MLST provides a powerful tool for the genotypic characterisation of bacteria at the inter-specific level, that is, it is able to show how related a group of species are (Maiden *et al.*, 1998).

| | Loci | | | | | | | ST |
|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | |
| Strain 1 | 1 | 3 | 1 | 1 | 2 | 4 | 1 | 1 |
| Strain 2 | 1 | 3 | 1 | 1 | 2 | 4 | 1 | 1 |
| Strain 3 | 1 | 3 | 1 | 1 | 2 | 4 | 1 | 1 |
| Strain 4 | 1 | 3 | 1 | 1 | 3 | 4 | 1 | 2 |
| Strain 5 | 2 | 1 | 2 | 3 | 1 | 1 | 2 | 3 |
| Strain 6 | 2 | 1 | 2 | 3 | 6 | 4 | 2 | 4 |
| Strain 7 | 2 | 1 | 2 | 3 | 1 | 1 | 2 | 3 |
| Strain 8 | 3 | 2 | 1 | 2 | 2 | 4 | 2 | 5 |
| Strain 9 | 3 | 2 | 1 | 2 | 2 | 4 | 2 | 5 |
| Strain 10 | 3 | 2 | 1 | 2 | 2 | 4 | 2 | 5 |

**Table 1** Allelic profiles and sequence types (ST).

MLST provides greater discriminatory power compared to other techniques (Kotetishvili *et al.*, 2003), given that a single base change may result in the assignment of a new allele, the probability of two unrelated isolates having the same

ST by chance is very small. By varying the number and type of genes used, MLST can be tweaked to address specific epidemiological questions. For example, most of the published schemes employ slowly evolving housekeeping genes, which can discriminate between species and major divisions within a species. Such data is useful for long term global epidemiology and the surveillance of major lineages. However, for the characterisation of strains corresponding to single localised outbreaks, such as in a single hospital ward or school, more discriminatory (variable) genes such as those encoding antigenic proteins may be employed. One of the most important benefits of MLST is that the technology used is generic and is therefore easily reproducible, and also allelic profiles can be shared online by uploading the information onto the MLST website (http://www.mlst.net/). The electronic portability of nucleotide sequence data means that results can be easily compared between laboratories anywhere in the world via the Internet, which is much more difficult with gel-based techniques such as PFGE or MLEE.

-Multi-locus sequence analysis

Multi-locus sequence analysis (MLSA) has been developed from MLST for the genotypic characterisation and species demarcation at the genus-wide level (Gevers *et al.*, 2005). Based on the typing protocol of MLST, each isolate is characterised on the basis of multiple protein-coding genes, and allelic profiles and STs are subsequently produced. Loci which can be used for analysis should be ubiquitous (at least to the taxon under study), present in single copy and in which recombination might confer a selective advantage (e.g. antigen-encoding genes), or closely linked genes across the genome. Within MLSA studies the nucleotide sequences for the multiple protein coding genes are concatenated to construct a phylogenetic tree. By concatenating the data, the fact that some genes are more diverse than others and are under varying levels of selective pressure, can be buffered for producing a more accurate phylogeny. For example, the 6 protein coding genes for *Neisseria meningitidis*, *N. lactamica* and *N. gonorrhoeae* each produce varying tree topologies, with in some cases more than one species clustering together (Fig. 6). When the sequence data is concatenated, three distinct clusters are observed, each corresponding to the three *Neisseria* species (Hanage *et al.*, 2005).

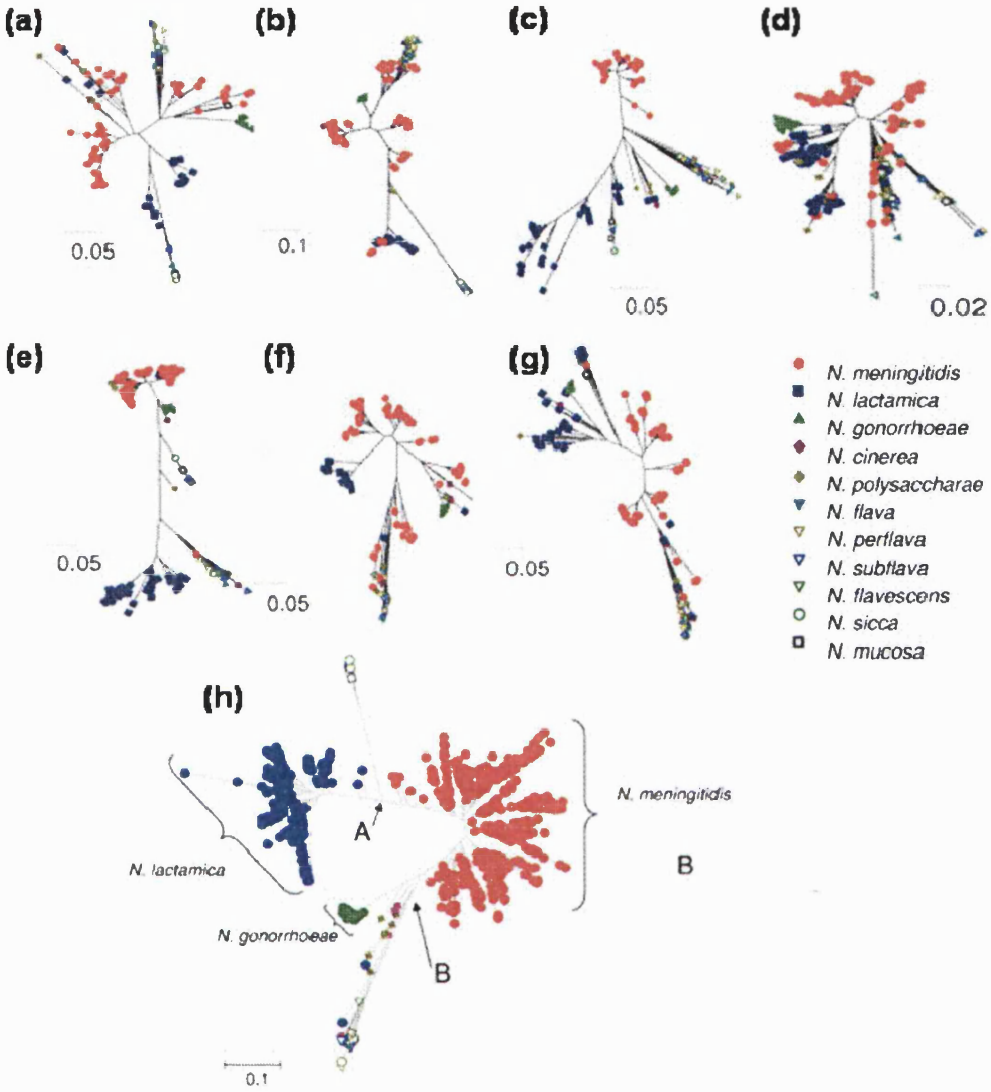



**Figure 6** Single locus trees fail to resolve species clusters. Individual phylogenetic trees for each of the MLST loci for *Neisseria* spp.; a) *abcZ*, b) *adk*, c) *aroE*, d) *fumC*, e) *gdh*, f) *pdhC* and g) *pgm*. Concatenation of the loci (h) reveals 3 distinct clades representing *N. lactamica*, *N. meningitidis* and *N. gonorrhoeae*. Figure from Hanage *et al.*, (2005).

## 1.7 Phylogenetic analysis

### -Distance methods

Distance methods for phylogenetic reconstruction involve the conversion of sequence alignments into a distance matrix of pairwise differences (or distances) between the sequences since the sequences diverged from their most recent common ancestor. The distance method then uses the matrix as the data for production of branching order and branch lengths. The simplest calculation of distance between two sequences involves counting the number of observed differences between the sequences (p-distance). This however does not account for multiple changes at one site. Therefore a variety of models have been developed to calculate the distances between DNA sequences, most of which differ only in how many parameters they attempt to include.

One of the simplest models of nucleotide substitution is that of Jukes & Cantor (Jukes and Cantor, 1969). In this model, it is assumed that nucleotide substitution occurs at any nucleotide site with equal frequency (so the expected frequencies of A, T, C & G will be equal to 0.25) and that at each site a nucleotide changes to one of the three remaining nucleotide frequencies with a probability of $\alpha$ . Therefore, the probability of change of a nucleotide to any of the three nucleotides is $r = 3\alpha$. Under this model, the expected number of nucleotide substitutions per site (d) between two sequences is:

$$d = -\frac{3}{4} \ln\left(1 - \frac{4}{3} p\right)$$

Other distance models, for example, Felsenstein (F81) (Felsenstein, 1981), Hasegawa, Kishino and Yano 1985 (HKY85) (Hasegawa *et al.*, 1985) and general time reversible (GTR) (Lanave *et al.*, 1984), add complexity to the distance calculation differing only in how many parameters they attempt to include. The Kimura's Two-Parameter model (K2P) (Kimura, 1980) estimates the number of nucleotide substitutions per site, taking into account that the rate of transitional nucleotide substitution ($\alpha$) is often higher than that of transversional nucleotide substitution ($2\beta$). Therefore the total substitution rate per site is $\alpha + 2\beta$. Under this

model, the expected number of nucleotide substitutions per site between two sequences is:

$$d = -\frac{1}{2}\ln(1-2P-Q) - \frac{1}{4}\ln(1-2Q)$$

With Kimura's model, the equilibrium frequency of each nucleotide is 0.25.

Distance methods can be made more parameter rich by using a gamma correction (this allows different nucleotide positions to evolve at different rates). Other parameters can be added to allow for sites that cannot change, for variables sites which can change and correction parameters to allow different substitution rates for each type of nucleotide change (Yang *et al.*, 1994). Distance methods are considered fast and therefore suitable for analysing large datasets (Tamura *et al.*, 2004).

-Bayesian inference of phylogeny

Bayesian inference of phylogeny is expressed as the posterior probability of a tree given the data (i.e. the probability that the tree is correct) (Huelsenbeck *et al.*, 2001). Bayesian methods are closely linked to likelihood models. They are similar in that the user postulates a model of evolution and the program searches for the best trees that are consistent with both the model and the data. The differences lie in the use of a prior distribution of a tree in Bayesian inference (i.e. maximum likelihood seeks the tree that maximises the probability of observing the data given the tree, whereas Bayesian analysis seeks the tree that maximises the probability of the tree given the model and the data). Bayesian analysis also searches for the best set of trees, whereas likelihood searches for a single tree (Hall, 2004). Calculating the posterior probability involves summation over all trees, integration over all possible combinations of branch lengths on a tree for each tree, and calculation of substitution model parameters (Huelsenbeck *et al.*, 2001). Therefore a Markov chain Monte Carlo (MCMC) numerical method is used to approximate the posterior probability more efficiently (Huelsenbeck *et al.*, 2001).

The MCMC method explores parameter space in a stepwise fashion. MrBayes uses the Metropolis-coupled MCMC method that searches with several chains, some of which are "heated". Heating is the proportional, exponential increase in the posterior probability of a step and allows chains to escape local optima (low peaks in tree-space where stepwise changes are likely not to find a tree with a higher posterior probability) (Huelsenbeck *et al.*, 2002). At specified intervals, chain states are swapped based on observed differences in likelihood. MCMC can be visualised as a set of independent searches that occasionally exchange information. Maximum likelihood searches the landscape of possible trees my moving from high points to higher points looking for optimal trees (Huelsenbeck *et al.*, 2002). If there is more than one hill, maximum likelihood can get trapped on a hill even if there are higher hills with more likely trees. MCMC on the other hand, allows the search to jump valleys to other high points in "tree-space" (Hall, 2004). Bayesian analysis can result in a program getting stuck on a sub-optimal tree, but the MrBayes program deals with this by the introduction of Metropolis-coupled MCMC, where several chains run in parallel (Huelsenbeck and Ronquist, 2001). In the MrBayes program, the chain begins with a tree (this tree can be randomly chosen or can be one that is likely to be a good description of the data) containing branch lengths and parameters for substitution and rate variation across sites (Hall, 2004).

The posterior probability for each parameter is based on the frequency with which the parameter values are observed. As the number of generations increases, the program closes in on a set of trees where the likelihoods are so close that rejecting change is a random choice. Here, the likelihoods are said to have converged on a stable value and have reached equilibrium (Hall, 2004). At this point, the MrBayes program samples trees according to their posterior probabilities. Trees sampled after equilibrium is reached are used to generate a consensus tree, built via Paup* version 4.0b10 (Swofford, 2000). Trees sampled before equilibrium are discarded as "burn-in" and omitted from the construction of the consensus tree. These are trees where the likelihood scores have not converged on a steady value (i.e. equilibrium) and therefore these are not optimal trees, and are removed.

## 1.8 Methods for the detection of recombination

### -Sawyer's Run Test

The Sawyer's Run Test looks for evidence of recombinational exchanges within a set of aligned sequences, by determining if regions of sequence pairs have more consecutive identical polymorphic sites in common than would be expected by chance (Drouin *et al.*, 1999). In this test, positions that are totally conserved are omitted (condensation) from the alignment and, for each sequence pair, a set of fragments is defined between successive sites where these two sequences differ. The lengths of all the fragments found between every pair-wise comparison are used to obtain two *P*-values. Firstly, the sum of the square of the condensed fragments (SSCF) is a measure of the distribution of the lengths of the condensed fragments, where shorter condensed fragments will contribute less to the SSCF value than will long fragments. Secondly, the maximum condensed fragment (MCF) is a measure of the distribution of the longest condensed fragment for all pairs (Drouin *et al.*, 1999; Sawyer, 1989). A gene conversion event results in a segment of bases in which the two sequences agree, therefore, an increase in SSCF & MCF values will occur. Provided the mutation rate is constant across sequences, SSCF & MCF are not significantly influenced by mutational hot and cold spots. SSUF and MUF are defined similarly, but with the omission of the condensation step, therefore the test is performed on all nucleotide sites (Sawyer, 1989).

Sawyer's Run Test generates artificial datasets based on a user defined number of site permutations. These permutations rearrange the sites which are assigned a class based on degeneracy, with the permutation of the sites being constrained so that sites of a given class can only be assigned to site positions of the same class in the original dataset. Therefore, in the presence of recombination SSCF and MCF values for the observed sequences are expected to be greater than those obtained for the randomly permuted sequences (Sawyer, 1989).

### -Population-scaled recombination rate (ρ)

LDhat is a program to analyse patterns of linkage disequilibrium within the history of a sample of genes (Hudson, 2001). The key parameter in determining the extent of linkage disequilibrium is the population-scaled recombination rate (ρ) $4N_e r$, where

$N_e$ is the effective population size and r is the per locus recombination rate per generation. The program implements Hudson's (Hudson, 2001) composite-likelihood estimate approach to estimate the population-scaled recombination rate conditioned on the estimate of mutation rate per site ($\theta = 4N_er$) from an approximate finite-sites version of the Watterson estimate. The estimate of $4N_er$ is taken as the value that has the highest composite-likelihood estimate (McVean *et al.*, 2002). If no recombination has occurred then $\rho = 0$.

-Minimum number of recombination events

The minimum number of recombination events ($R_M$) in the history of a sample is based on the four-gamete test, which infers a recombination event between pairs of diallelic loci at which all four possible gametic types are present. Therefore, the sample size must be at least four and only segregating sites are compared. Under the assumption that overlapping regions signal the same recombination event, the $R_M$ is the number of non-overlapping intervals from the original sample, each one indicating different recombination events (Hudson and Kaplan, 1985; Myers and Griffiths, 2003). $R_M$ has the advantage of being very fast computationally, however it is not optimal; the $R_M$ will underestimate the total number of recombination events.

-Splits decomposition

Splits decomposition was developed to address the problem of conflicting phylogenetic signals within phylogenetic trees (Bandelt and Dress, 1992). It identifies incompatible partitions within sequence data, which commonly arise from recombination, by representing the relationship between sequences as a network rather than a tree (Huson and Bryant, 2006). In the presence of recombination, for example, the sequence data will show conflicting phylogenetic signals because different partitions may support different phylogenies. The advantage of using a splits network rather than a phylogenetic tree to represent evolutionary relationships between a given set of taxa, is that a tree presumes the underlying process to be either bifurcating or multifurcating, whereas a split network does not. Within a split network, every edge is associated with a split of the taxa, but there may be a number of parallel edges associated with each split. The length of an edge in the network is

proportional to the weight of the associated split. This is analogous to the length of a branch in a phylogenetic tree (Huson and Bryant, 2006). A fit close to 100% indicates true representation, whereas lower fits indicate a reduction in confidence.

-Classification index

The Classification index was developed to assess the extent of genetic differentiation between two populations. This index was originally used by Jolley *et al.*, (2005) in their comparison of disease and carriage *N. meningitidis* populations, and is calculated according to the equation:

$$D = 1 - \sum_{i} \frac{p_{i1}p_{i2}}{\bar{p}}$$

Where, $P_{ij}$ is the frequency of allele or ST $_i$ in population $_j$ (either 1 or 2), and the denominator $p$ (bar) is the average frequency of $p$ across the two populations. Similar to $F_{ST}$ where the populations are identical in terms of frequency the statistic is zero, where there is no overlap the statistic is 1. In cases where only two alleles are present at a locus the statistic is exactly equivalent to $F_{ST}$, but is more sensitive in cases where there are >2 alleles per locus as it considers the frequency of each allele rather than relying on a summary statistic such as homozygosity (Jolley *et al.*, 2005). The significance of the statistic was gauged by resampling 1000 randomised data sets from the pooled data according to the sizes of the real populations. This was carried out using a PERL script written by Dr. E. Feil.

-Tests of neutrality

In the study of molecular population genetics, the neutral theory proposed by Kimura (Kimura, 1968) plays an important role. In this theory, genetic variation at the molecular level is considered largely neutral, and the extent of variation is determined primarily by the mutation rate and the effective population size (Kimura and Crow, 1964). Therefore, it is possible to test the hypothesis of neutral evolution by comparing the observed and predicted amounts of genetic variation. If the discrepancy between the observed and predicted amounts is large, one may invoke

selection or a demographic change in the population. For example, negative values reflect an excess of rare polymorphisms in a population, which is consistent with either positive selection or an increase in population size. Positive values indicate an excess of intermediate-frequency alleles in a population and can result from either balancing selection or population bottlenecks.

- Wrights inbreeding coefficient

Wrights inbreeding coefficient ($F_{ST}$) compares the level of genetic variation within two or more sub-populations relative to the overall total population (Wright, 1951), and is estimated according to the equation:

$$F_{ST} = \frac{(T-S)}{T}$$

Where, $T$ is the average difference between pairs of alleles (or allele frequencies) drawn at random from the total population, and $S$ is the average difference between pairs of alleles (or allele frequencies), both drawn at random from the same subpopulation. If genetic exchange in the total population is random, then $T$ and S are expected to have the same value and $F_{ST}$ will be zero. However, if genetic exchange between sub-populations is non-random (i.e. they are partially isolated), genetic variation in each sub-population will be less than variation in the total population, and $F_{ST}$ will increase to a maximum of 1.0 (Bamshad, 2004).

-Index of association

This statistical test attempts to measure the extent of linkage equilibrium within a population by quantifying the amount of recombination among a set of sequences and detecting association between alleles at different loci (Smith *et al.*, 1993). The Index of Association ($I_A$) is calculated as follows:

$$I_A = (V_O/V_E) - 1$$

$V_0$ is the observed variance of K and $V_E$ is the expected variance of K, where K is the number of loci at which two individuals differ. If there is linkage equilibrium

because of frequent recombination events, the expected value of $I_A$ is zero. Clonal populations are identified by an $I_A$ value that differs significantly from zero.

-Homoplasy test

The Homoplasy Test was designed to test recombination in nucleotide sequence data sets derived from closely related organisms and by compatibility matrices, which can reveal reticulate evolution (Maynard Smith and Smith, 1998). If the same mutation occurs twice independently in the ancestry of a set of sequences, this is called a homoplasy. One can then calculate the homoplasy ratio, *H*, is a number whose expectation is 0 if the population is clonal and 1.0 if it is in linkage equilibrium. For simplicity, this analysis is confined to synonymous sites only.

## 1.9 eBURST

Based on the BURST (Based Upon Related Sequence Types) algorithm developed by Edward Feil (University of Bath), eBURST is currently the easiest and most efficient way of visualising the relationships between closely-related isolates of a bacterial species or population (Spratt *et al.*, 2004). eBURST assumes an ancestral (founding) genotype increases in frequency in a population, and while doing so, diversifies to produce a cluster of closely-related genotypes which are all descended from the founding genotype (Fig. 7a). A cluster of closely related genotypes is referred to as a "clonal complex" (CC). The procedure was developed for use with the data produced by MLST (STs and their allelic profiles) but can be used with some other molecular typing methods that define isolates as strings of integers.

To use eBURST, input data in the form of the allelic profiles and STs created from MLST, is loaded onto the eBURST website http://eburst.mlst.net/. The eBURST algorithm divides the input data into "groups" of STs that have defined levels of similarity in allelic profiles (Spratt *et al.*, 2004). A "group" is defined according the number of loci the group has in common, for example, if the input allelic profiles consist of 7 housekeeping genes, then the most stringent (conservative) "group" will be one in which every member shares identical alleles at 6 of the 7 loci with at least one other member of the group. The stringency of the grouping can be relaxed by

decreasing the number of alleles each of the STs has in common. The fewer loci that are used, the less chance there is for the group to belong to a single clonal complex. By assigning zero loci in common within the group, a "population snap shot" can be visualised detailing the relationships between each ST within the entire population.

Once the number of loci has been chosen for stringency (for descriptive purposes 5 out of 7 loci in common), eBURST produces an output diagram which visualises the groups and automatically links STs which are only different at one locus (single locus variants; SLVs (Fig. 7b). Each ST is represented by a black circle with the exception of the founding genotype which is represented by a blue circle and "subgroup founders" which are yellow. The size of the circle reflects how many isolates have that particular ST. The primary founder of a group is defined as the ST that differs from the largest number of other STs in the group at only one single locus. Multiple founding genotypes may be found within a population. Occasionally, a founding genotype may be misassigned by the eBURST algorithm due to sampling bias; in this case the user can re-assign the founder ST which will then be highlighted by a red circle. STs which have changed over time and have allelic profiles that differ from the founder at only one locus of the seven are called single-locus variants (SLVs); these are highlighted by black lines on the eBURST diagram. If an SLV has diversified further to produce an ST which differs at two of the seven loci from the founder, this ST is called a double-locus variant (DLV). If the intermediate SLV is not present in the sample, then DLVs can be linked directly (shown in blue). For further details, see the readme file at http://eburst.mlst.net.

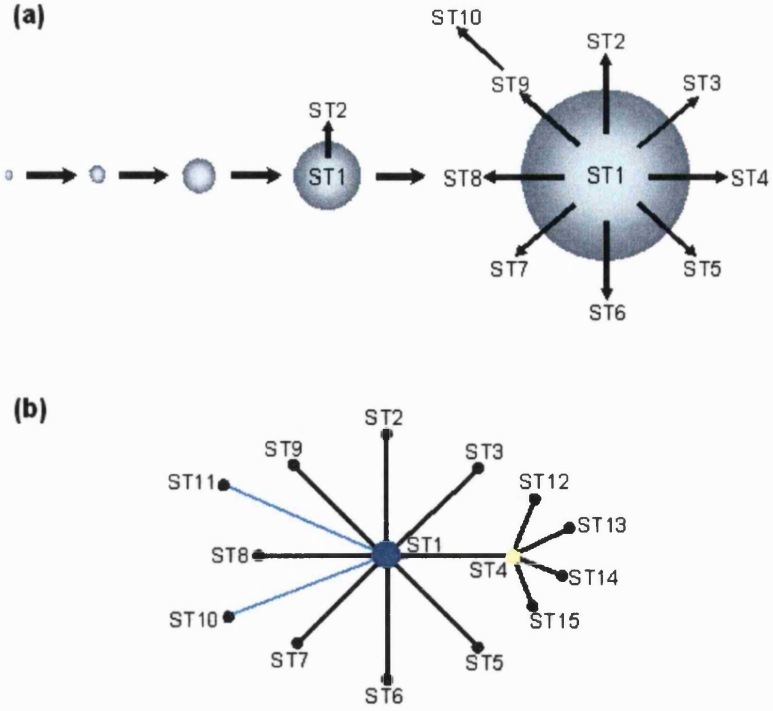



**Figure 7** Bacterial relatedness visualised by eBURST. As a founding genotype increases in frequency within a population (a) it can diversify to produce a cluster of closely-related genotypes, i.e. a clonal complex. This is visualised in eBURST in section (b) where ST1, the founder ST (blue circle), has diversified to produce 8 SLVs (black lines), one of which is a sub-group founder (yellow), which has further diversified to produce 4 SLVs. ST1 has two DLVs (blue lines).

## 1.10 Aims of this thesis

The aim of this thesis was to answer some outstanding questions in the population structure of bacterial host interactions, particularly those relating to host adaptation and HGT. Therefore, each chapter deals with one (or more) aspect(s) of bacterial/host evolution by studying either natural (*Wolbachia* spp. and *Vibrio* spp.) or experimental (*Staphylococcus aureus*) bacteria / host interactions in order to understand how these interactions shape the genetic variation and population structure of the bacteria.

In Chapter Three, the evolution of the natural population of the bacterial endosymbiont *Wolbachia* spp. is examined using MLSA. *Wolbachia* was isolated from a range of arthropod hosts, across the globe in order to address the following:

- Determine the utility of MLSA in *Wolbachia* evolution and phylogenetics;
- Determine the degree of *Wolbachia* spp. host congruence between multiple arthropod Genera;
- Compare the importance of small-scale and large-scale geographical distance in generating diversity;
- Establish the utility of alternative loci for *Wolbachia* supergroup phylogenies;
- Estimate the level of HGT within the global population;

Chapter Four extends the analysis of Chapter Three by examining the molecular evolution of *Wolbachia* spp. within a single Family of arthropods; the spider mites. This study aims to address the following:

- Determine the degree of host congruence within the spider mite study population;
- Compare the importance of small-scale and large-scale geographical distance in generating diversity in *Wolbachia* recovered from a fairly sedentary host, and to examine how this varies according to the resolution provided by different phylogenetic markers;
- Determine if *Wolbachia* isolated from spider mites is restricted to *Wolbachia* supergroup B;
- Provide further evidence of horizontal transmission within *Wolbachia* spp.

The marine *Vibrio* spp. population is investigated in Chapter Five and was conducted across two sites at the Gower Peninsula, Wales, over a period of 17 months, sampling a number of eukaryotic host species. This study aims to address the following:

- Determine the extent of seasonal variation within the *Vibrio* marine population;
- Determine the degree of local and host adaptation, and the level of HGT in both;

- Test the utility of *recA*, the traditional *Vibrio* species assignment locus, against multiple protein coding loci and MLSA.

Establishing the extent of local adaptation on a global scale is crucial for the development of the proper understanding of the genetic structure of any bacterial population. Chapter Six addresses the fundamental questions concerning patterns of migration and gene flow within the *Vibrio* genus, and marine bacteria in general, by the comparison of *Vibrio* isolates recovered at the same time, but from opposite sides of the world. This study aims to address the following:

- Compare the importance of small-scale and large-scale geographical distance in generating diversity;
- Provide further evidence for recombination and selection over global scales;
- Examine the hypothesis, *"Everything is everywhere, but the environment selects"* within the *Vibrio* study population.

Chapter Seven aims to develop an *in vivo* model of *Staphylococcus aureus* infection based on the insect larvae *Manduca sexta* (the Tobacco Horn Worm), examining differences in virulence potential between natural isolates and to draw comparisons with the virulence gene profiles and epidemiological origin of these isolates. The study is based upon the findings of Peacock *et al.*, (2002) who established 7 out of 33 virulence genes were significantly more likely to be present in strains from invasive disease than in strains from asymptomatic carriage. This study aims to address the following:

- Establish an insect *in vivo* model of infection;
- Identify, *in vivo*, reproducible differences in virulence between invasive and asymptomatic *S. aureus* isolates;
- Identify any specific genetic factors (such as the global gene regulator), which is associated with virulence in *S. aureus* and explore the relevance of this association.

I conclude in Chapter Eight by reviewing the major findings of this thesis and the implications for bacterial host interactions.

# CHAPTER TWO

## MATERIALS AND METHODS

# 2. MATERIALS AND METHODS

## 2.1 Bacterial strains

### *Wolbachia* spp.

*Wolbachia* spp. used in this study were isolated from numerous Orders of arthropods collected from across the globe, all within the past 5 years. *Wolbachia* spp. isolated from *Bryobia kissophila, Bryobia sarothamni, Bryobia praetiosa* and *Tetranychus urticae* were kindly donated by Dr. Hans Breeuwer (University of Amsterdam). The remaining invertebrates for the study were kindly donated by Dr. Robert Butcher (then of University of Bath) and Dr. Tim Karr (University of Bath).

### *Vibrio* spp.

*Vibrio* spp. were isolated over a 17 month period from two sites at the Gower Peninsula, Wales. The target host organism of the study was *Actinia equina* (the Beadlet sea anemone), although isolates were also recovered from periwinkles (*Littorina* spp.), limpets (*Patella vulgata*), common shore crabs (*Carcinus maenas*), mussels (*Mytilus edulis*) and sea water. Sequence data collected from a similar environmental *Vibrio* spp. study conducted in New Zealand were kindly made available by Professor Paul Rainey (University of Auckland). These strains were all recovered from sea water or *A. tenebrosa.*

Further details of isolates used are given within the Methods section of the relevant chapter.

## 2.2 Preparation and storage of cell and DNA stocks

### *Wolbachia* spp.

Invertebrates from which *Wolbachia* was isolated were stored at -20°C in 95 % ethanol.

Genomic DNA was extracted using the Nucleospin Kit (Machery-Nagel, Germany) following the manufacturers instructions. Briefly, this procedure involved the following steps:

- Insect samples were homogenised using a small pellet pestle in a 1.5 ml eppendorf tube.
- Samples were lysed with proteinase K.
- Samples were added to Nucleospin binding columns to allow the genomic DNA to bind to the membrane and were then washed.
- Genomic DNA was eluted from the membranes using the elution buffer supplied with the kit.

DNA stocks were then stored at -20°C.

***Vibrio* spp.**

-Growth media preparation

Thiosulphate Citrate Bile Sucrose agar, TCBS, (Oxoid, UK) was used for the isolation and selective cultivation of *Vibrio* spp. Media was prepared as follows:

- 20 g of TCBS dissolved in 500 ml sterile distilled water.
- Media was boiled in a microwave (media cannot be autoclaved), firstly for 2 minutes at full power (media should then be mixed gently) and then for a further 2.5 minutes on full power (media should be mixed gently each time it begins to boil).
- Media was left to cool and once at a suitable temperature was poured into 60 mm plastic petri dishes.
- Petri dishes should be stored at 4°C "agar side down" due to the instability of the media.

Environmental isolates were grown at 26 – 28°C for 12 – 24 hours.

-Isolation of samples

*Vibrio* species were recovered using different techniques according to the host organism/source:

- Sea anemones
  - *Vibrio* species were recovered by swabbing *A. equina/A. tenebrosa* individuals gently at their actinopharynx at low tide (Fig. 1). Sea anemones were not removed from their location; however their

positions were recorded in addition to the air temperature. Single swabs were streaked across TCBS media.

- Hard-shelled eukaryotes
  - The periostracum of periwinkles, limpets, crabs & mussels were removed and the inner body homogenised thoroughly before the homogenate was spread-plated onto TCBS media.

In addition to TCBS media, 0.5% (w/v) sodium chloride alkaline peptone water (Oxoid, UK) was used for the potential enrichment any of *V. cholerae* cells present. Each bijou (as supplied by the manufacturer) contained approximately 5ml of alkaline peptone water. *Vibrio* species were isolated either by immersing the used cotton swab into the broth or by the placing of the marine host into the broth without homogenisation. Isolates were grown at 26-28°C for 12-24 hours or until turbid.

To ensure the isolation of pure, single colonies (i.e. a single *Vibrio* species):

- TCBS
  - Single colonies were picked at random from each plate (approximately 5 colonies) and subcultured onto new TCBS media.
- Alkaline peptone water
  - One drop of broth was spread-plated onto TCBS media and incubated overnight. From this plate, single colonies were picked at random and subcultured onto new TCBS media.

**Figure 1** Delicate swabbing of *A. equina* at the actinopharynx.

-Preparation of frozen cell stocks

Single colonies were resuspended in 1 ml LB broth (Fisher Bioreagents, UK) with 15% (w/v) glycerol (Sigma, UK) and stored at -80°C providing frozen cell stocks.

-Genomic DNA extraction

Genomic DNA was extracted using the Nucleospin Kit (Machery-Nagel, Germany) following the manufacturers instructions. Briefly, this procedure involved the following steps:

- A single colony was picked from TCBS media and resuspended.
- Samples were lysed with proteinase K.
- Samples were added to Nucleospin binding columns to allow the genomic DNA to bind to the membrane and were then washed.
- Genomic DNA was eluted from the membranes using the elution buffer supplied with the kit.

DNA stocks were then stored at -20°C. All stocks were catalogued and thermostably labelled for traceability and accurate identification.

## 2.3 Primer design

Three methods were used in the design of MLSA primers:

- By Eye

  Primers designed by eye met the following criteria:

  - At least 30 nucleotides from the beginning/end of gene.
  - 18 - 27 nucleotides in length.
  - GC content of 40-60 %.
  - Annealing temperatures in the range of 50-65 °C. In addition, forward and reverse primers should anneal at approximately the same temperature (allowing perhaps 3 or 4 °C of difference between them).
  - Do not contain palindromes.
  - Do not contain sequences of nucleotides that would allow one primer molecule to anneal to itself or to the other primer used in the PCR reactions (primer dimer formation).
  - Do not end on the variable 3$^{rd}$ position of a codon.

- CODEHOP

  The program CODEHOP was used to design primers from multiple protein sequence alignments (Rose *et al.*, 1998). The program is intended for cases where the protein sequences are distant from each other and degenerate primers are needed. CODEHOP primers were designed as follows:

  - They are degenerate at the 3' core region (approximately 11-12 bp in length across four codons of highly conserved amino acids) and non-degenerate at the 5' consensus clamp region (approximately 20 – 30 bp in length, dependent on the desired annealing temperature).

- The hybrid structure (5' consensus and 3' degenerate) of CODEHOP primers allow the PCR amplification to be specific during the early cycles from the original source DNA and selective during the late cycles from the PCR synthesized products.
- File inputs for the CODEHOP program must be aligned and in FASTA format.

- Basic Local Alignment Search Tool (BLAST)

  Using the Basic Local Alignment Search Tool algorithm, implemented at the NCBI website (http://www.ncbi.nlm.nih.gov/BLAST), Blastn was used to locate conserved regions of nucleotides which could be used as putative primer sites. Primers were designed as follows:
  - Genome specific nucleotide sequence data for the gene in question was entered into Blastn. Two formatting options were altered:
    - Results were limited by selecting only bacteria.
    - The alignment view was set to "query-anchored with identities" (this allows conserved regions to be viewed easily across the Blast query hits).
    - Once conserved regions have been identified which are a suitable distance apart for PCR and DNA sequencing, primers can be designed, building in degenerate bases when needed according to the alignments.

Furthermore, a number of "universal" primers were tested to determine their suitability within a MLSA study (Santos and Ochman, 2004). Santos & Ochman developed primer sets that targeted 10 genes spanning an array of genomic locations and functional categories. Primers were designed using the program CODEHOP employing the following parameters:

- Annealing temperature of at least 55°C.
- Equal codon usage.
- Sequence degeneracy of $\leq 512$.

Detailed information regarding specific primers used within the MLSA studies can be found in:

- Chapter 3 and Appendix A for *Wolbachia* spp.
- Chapter 5 and Appendix C for *Vibrio* spp.

## 2.4 Methods for polymerase chain reaction and sequencing

### -DNA amplification: polymerase chain reaction (PCR)

PCR reactions were run in 46 well microtitre PCR plates (ABgene, UK), and contained:

***Wolbachia* spp.**

| | |
|---|---|
| 22.5 µl | Gold PCR Master Mix (Biogene, UK) |
| | 1.25 u *Taq* polymerase |
| | 200 µl dNTPs |
| | 1.5 mM $MgCl_2$ |
| | 5% Sucrose |
| | 1 mM Gold Dye |
| | 75 mM Tri-HCl (pH 8.8 at 25°C) |
| | 20 mM $(NH_4)_2SO_4$ |
| | 0.01% (v/v) Tween20 |
| 1 µl | Forward primer (10 pmol-1, MWG Biotech, Germany) |
| 1 µl | Reverse primer (10 pmol-1, MWG Biotech, Germany) |
| 1.8 µl | 25 mM $MgCl_2$ (final $[MgCl_2]$ = 3.5 mM, Biogene, UK) |
| 1 µl | DNA |

***Vibrio* spp.**

All PCR reactions contain the following reagents:

| | |
|---|---|
| 12.5 µl | ReddyMix™ PCR Master Mix (ABgene, UK) |
| | 1.25 u Thermoprime plus DNA Polymerase |
| | 75 mM Tris-HCl (pH 8.8 at 25°C) |
| | 20 mM $(NH4)_2SO_4$ |
| | 1.5 mM $MgCl_2$ |
| | 0.01% (v/v) Tween® 20 |

| | |
|---|---|
| | 0.2 mM each of dATP, dCTP, dGTP, dTTP |
| | Precipitant and red dye for electrophoresis |
| 9.5 µl | distilled $H_2O$ |
| 1 µl | Forward primer (10 pmol-1, MWG Biotech, Germany) |
| 1 µl | Reverse primer (10 pmol-1, MWG Biotech, Germany) |
| 1 µl | DNA |

The following standard PCR conditions were used for all PCR reactions, unless otherwise stated. Annealing temperatures and extension times for each reaction were adjusted according to primer pairs.

-PCR conditions:

| | | | |
|---|---|---|---|
| Initial denaturation | | 3 mins | 94°C |
| 34 cycles: | denaturation | 1 min | 94°C |
| | annealing (primer specific) | --- | --- |
| | extension (primer specific) | --- | 72°C |
| Final extension | | 10 mins | 72°C |

All primer sequences, annealing temperatures and extension times can be found in each individual chapter.

-Electrophoresis of PCR product

The presence of amplicons was detected by gel electrophoresis stained with Ethidium Bromide.

- Agarose gel preparation

  A 0.8% agarose gel (10ml TBE Buffer 10X (Promega, Wisconsin, USA), 90 ml $dH_2O$, 0.8g agarose, 1 µl Ethidium Bromide (10 mg/ml)) was run in 1% TBE buffer (100 ml TBE Buffer 10X, 90 ml $dH_2O$).

- Loading sample

  5 µl dH2O, 2 µl loading dye (blue/orange 6X, Promega, Wisconsin, USA) and 5 µl 1Kb DNA ladder (Promega, USA) was mixed and loaded into a well to provide accurate calibration of amplicon size. 16-20 % of amplified PCR product was directly loaded into the wells of the agarose gel. The samples

were electrophoresed for 40-50 minutes at 85 volts and visualised on the "Uvidoc" UV transilluminator.

-Purification of PCR products

Amplicons were precipitated by the addition of 52 µl of NaOAc/EtOH solution to samples in microtitre plates (2 µl 3M NaOAc, 50 µl 95% EtOH). The PCR plates were sealed, vortexed, centrifuged to at least 1000 rpm for 10 seconds and then incubated at -20°C for 1 hour. After incubation, plated were centrifuged at 3500 rpm, 4°C for 1 hour. Immediately following this, plates were inverted onto fresh blue roll and spun at 500 rpm for 1 minute in order to remove residual ethanol from the wells. The amplicon was washed by the addition of 150 µl of 70% ethanol. The plates were resealed and spun at 3500 rpm for 30 minutes. Following this, supernatant was discarded by inversion onto fresh blue roll and the inverted plates centrifuged at 750 rpm for 1 minute. Plates were air dried for 20 minutes to ensure complete removal of residual ethanol. The purified amplicon was then resuspended in 15 µl distilled $H_2O$.

In cases where small numbers of PCR amplicons required purification, the QIAquick PCR purification kit (Qiagen, Crawley, UK) was used according to manufacturers instructions.

-Sequencing of purified amplicons

Sequencing reactions were performed in microtitre PCR plates, with each reaction containing 2 µl of precipitated purified amplicon, 1 µl primer (1 $pmol^{-1}$), forward or reverse (2 wells per isolate – forward/reverse), and 2 µl *Taq* FS-Big Dye (Version 3, Applied Biosystems, UK). The plate was then sealed and centrifuged to at least 1000 rpm for 10 seconds. The plates were then placed on the DNA engine. The sequencing PCR was performed with an initial 10 seconds denaturation at 96°C, followed by 24 cycles at 50°C, extension at 60°C for 2 minutes, and then 4°C forever.

-Precipitation of sequence cycling products

The final precipitation of the sequence cycling products was carried out by the addition of 52 µl of NaOAc/EtOH solution (2 µl 3M NaOAc, 50 µl 95% EtOH).

Plates were sealed, vortexed, centrifuged to at least 1000 rpm for 10 seconds and then incubated at -20°C for 1 hour. After incubation, plates were centrifuged at 3500 rpm, 4°C for 1 hour. Immediately after centrifugation, plates were inverted onto fresh blue roll and spun at 500 rpm for 1 minute in order to remove residual ethanol from the wells. The pellet was washed by the addition of 150 µl of 70% ethanol. The plates were resealed and spun at 3500 rpm for 30 minutes. Following this, supernatant was discarded by inversion onto fresh blue roll and the inverted plates centrifuged at 750 rpm for 1 minute. Plates were air dried for 20 minutes to ensure complete removal of residual ethanol. The plates were then sealed and stored at -20°C until ready for loading on the ABI Prism 377 DNA sequencer.

## 2.5 Nucleotide sequence analysis

Nucleotide sequence trace files were assembled and edited using SEQMAN, DNAstar (Lasergene Inc, WI, USA) and BioEdit (Hall, 1999). The coding strand was determined using the translation package EDITSEQ, DNAstar. ClustalX (Thompson *et al.*, 1997) was used to create multiple alignments. Fragment size and conserved start and stop sequences for genes and gene fragments were identified for individual genes using ClustalX. Genes with regions of hypervariability were aligned using the tool MUSCLE (Edgar, 2004) and a wraparound TCL script kindly written by Prof. Laurence Hurst. This is a powerful alignment tool which back translates nucleotide sequence data to ensure it is in the correct coding nucleotide alignment. In addition, is it able to align multiple sequences in a fraction of the time taken by ClustalX. Therefore, MUSCLE is much more useful that ClustalX due to ClustalX's difficulty in aligning very variable sequences which contain many gaps and slow processing speeds.

## 2.6 Phylogenetic analysis

### Distance methods

The Kimura's Two-Parameter model (K2P) (Kimura, 1980) was implemented using MEGA version 3.1.1 (Kumar *et al.*, 2004) to create Neighbour-Joining phylogenetic trees. 1000 bootstrap replicates were performed.

MEGA version 3.1.1 was also used to determine the levels of synonymous and nonsynonymous mutations, the percentage of variable sites and the average pairwise diversity ($\pi$) for each locus.

Bayesian inference of phylogeny

Before nucleotide sequence data could be entered into the MrBayes (Huelsenbeck and Ronquist, 2001) program (version 3.1.1) it was exported from BioEdit into NEXUS format, and then a "Bayes block" was added; this contains all the necessary information & parameters for the MrBayes program to run.

Important parameters used are as follows:

- number of generations = see individual chapters for further details
- number of chains = 4
- sampling frequency = 1000
- print frequency = 100
- burnin = 20%.

Paup* version 4.0b10 (Swofford, 2000) was used to produce 50% majority rule consensus trees with data generated by MrBayes. Branch length data was incorporated in the tree building process.

## 2.7 Tests for recombination

Population-scaled recombination rate ($\rho$)

The program LDhat version 2.0 was used to analyse patterns of linkage disequilibrium within each of the loci (McVean *et al.*, 2002). Data was entered in FASTA format and linkage parameters were set with a minimum of 0 and a maximum of 100.

DnaSp

To assess the population mutation rate (Watterson's θ), recombination rate (Hudson's R), extent of neutral evolution (Tajima's D, Fu & Li's F* & D* statistics) and the genetic variation between populations ($F_{ST}$), FASTA formatted nucleotide sequence data was used within DnaSp version 4.10 (Rozas *et al.*, 2003).

START

The program START version 1.05 (Jolley *et al.*, 2001) is a selection of tools used for the analysis of MLST data. To determine values for Sawyer's Run Tests (10,000 permutations), Index of Association ($I_A$) and the Homoplasy Test, FASTA formatted nucleotide sequence data and allelic profiles were entered into START.

Classification Index

To assess the extent of genetic differentiation between two populations, the Classification Index was used. This was calculated by the method outlined in (Jolley *et al.*, 2001). A Perl script written by Dr. E. Feil was used to determine the significance of the result.

Splits decomposition

Splits decomposition analysis was performed using Splits Tree version 4.1 (Huson and Bryant, 2006) to examine the extent to which the data are compatible with a bifurcating tree or a network. Data was entered in FASTA format.

## 2.8 Other methods of analysis

eBURST

eBURST version 3 (Feil *et al.*, 2004) was used to display the relationships between closely related isolates of the UK and NZ populations. Before eBURST could be used, FASTA formatted nucleotide sequence data was entered into the web-based program NRDB (non-redundant databases). Once allelic profiles for each gene were generated, and STs assigned, the data was entered into eBURST. In addition, nucleotide sequence data was translated using MEGA (Kumar *et al.*, 2004) and allelic profiles assigned using NRDB. The translated STs were also entered into eBURST by way of comparison. The use of translated sequences for eBURST is

novel, and may prove a powerful means by which this approach, which is designed for the analysis of different strains of a single species, may reveal patterns of clonal expansion from strains encompassing multiple species. To differentiate between nucleotide and translated, STs and clonal complexes, the prefix "nt-" and "aa-" will be used respectively.

Datamonkey

Datamonkey is a web interface (http://www.datamonkey.org/) to a suite of maximum likelihood-based tools for identification of sites subject to positive or negative selection (Pond and Frost, 2005). Data was entered in FASTA format.

# CHAPTER THREE

## CHARACTERISATION OF THE NATURAL POPULATION OF *WOLBACHIA* SPP. BY MULTI-LOCUS SEQUENCE ANALYSIS

## 3.1 INTRODUCTION

### 3.1.1 *Wolbachia* spp.

The Gram-negative genus *Wolbachia* encompasses obligately intracellular bacteria in the Anaplasmatacea family of α-Proteobacteria (Werren, 1997a), and are most closely related to the arthropod-transmitted bacteria *Ehrlichia, Anaplasma, Neorickettsia* and *Rickettsia* (Dumler *et al.*, 2001) (Fig. 1). First discovered by Hertig and Wolbach (Hertig and Wolbach, 1924) in the reproductive tissues of the mosquito *Culex pipiens* in 1924, this *Rickettsia*-like bacterium was subsequently named *Wolbachia pipientis* in honour of Burt Wolbach (Hertig, 1936).

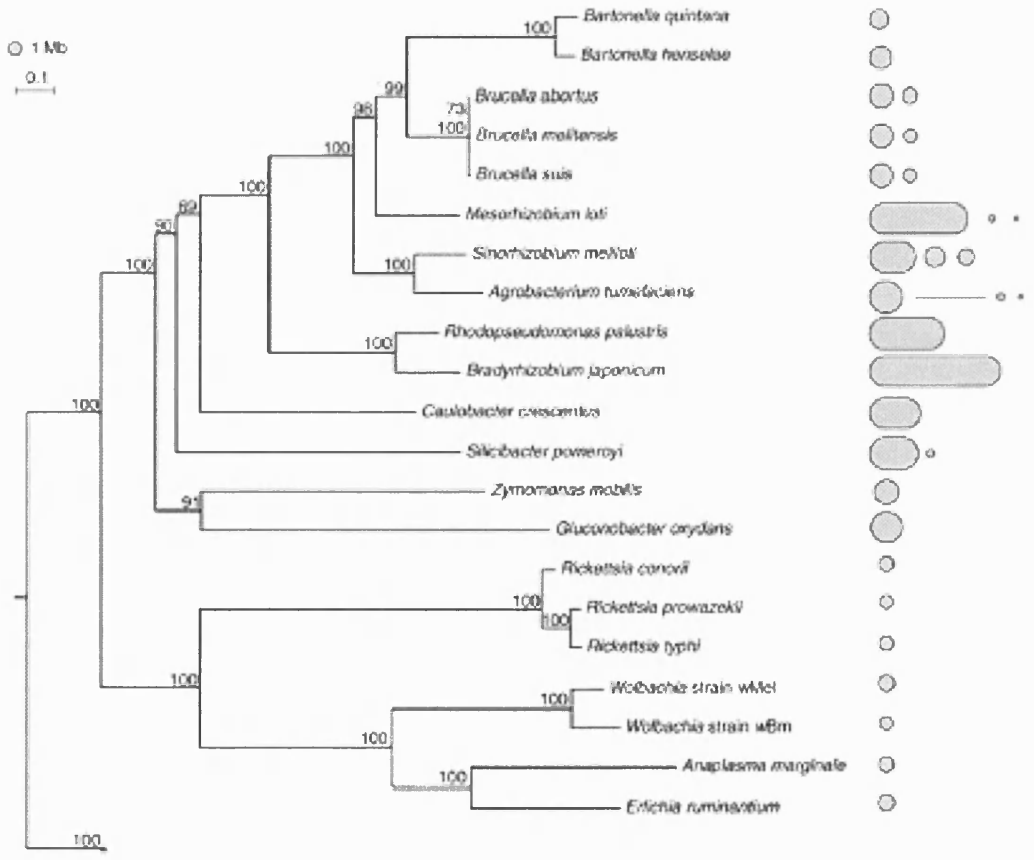



**Figure 1** Phylogenetic tree reconstruction for complete genome α-Proteobacterial species. Numbers refer to the statistical support for the divergence nodes (100 replicates). Graphics represent replicon sizes and architectures. Figure from Sallstrom and Andersson (2005).

*Wolbachia* are dimorphic microorganisms, with very small irregularly formed rod-like (0.5 - 1.3 μm in length) and coccoid forms (0.25 - 0.5 μm in diameter) (Stouthamer *et al.*, 1999; Woo Oh *et al.*, 2000). The bacteria are typically present in a vacuole, enveloped by three layers of membranes (Fig. 2). The outer layer is of host origin, which is followed by the outer cell wall of the bacteria. The innermost layer consists of the plasma membrane of bacteria (Louis and Nigro, 1989; Woo Oh *et al.*, 2000). Intracellular bacteria are often surrounded by multiple membranes, suggesting these structures may play a role in the eukaryotic host's control over the microorganism. The number of *Wolbachia* per host varies substantially, from as little as 250 cfu per egg observed within the genus *Trichogramma*, to 36.5 x $10^6$ cfu within male *Drosophila simulans* from Riverside (Bourtzis *et al.*, 1996; Stouthamer and Werren, 1993a). It is not possible to culture *Wolbachia* spp., as with other bacteria, in a cell-free medium. This has hindered *Wolbachia* research substantially leading to a bias towards infections that occur in host insects that are easily reared. One possible solution to this problem is the use of tissue culture to maintain these symbionts outside of the invertebrate host, and to date a number of *Wolbachia* spp. from various hosts (e.g. *Aedes albopictus, Spodoptera frugiperda* and *Drosophila melanogaster* (Dobson *et al.*, 2002)) have been grown and maintained in the laboratory.

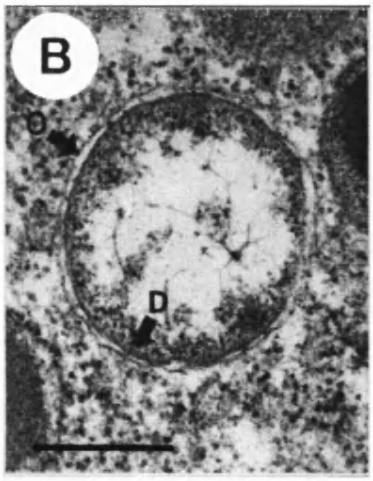



**Figure 2** Ultrastructure of endosymbiotic *Wolbachia* bacteria. High magnification of the bacteria shows the general characteristics of *Wolbachia*, including the double membrane (D) of the cell wall and an outer layer (O) of host origin. The cytoplasm consists of ribosomes and nucleic acid fibrils. Bar = 0.5 μm. (Woo Oh *et al.*, 2000).

### 3.1.2 *Wolbachia* nomenclature

The naming of *Wolbachia* strains is confusing and complicated, and urgently requires standardisation. Whereas some species names currently in use, e.g. *W. postica, W. popcorn* and *W. trichogramme*, are not formally recognised, other species such as *W. pipientis* and *W. persica* are. As with other bacteria (see Chapter 1), the application of the biological species concept is problematic for *Wolbachia* as it is unclear to what extent species clusters reflect adaptation to specific hosts. Although, gene flow might be expected to be negligible in *Wolbachia*, as these are bacteria are not thought to exist outside the insect host, a recent study reported pervasive gene exchange (Baldo *et al.*, 2005a).

Rousset and de Stordeur (Rousset and de Stordeur, 1994) first proposed a nomenclature system based on the insect host name or as the abbreviation style *w*Host, where *w* stands for *Wolbachia.* Based on this convention, *Wolbachia* strains that infect natural populations of *Drosophila simulans* in Riverside, California and Hawaii were named as *w*Ri and *w*Ha. Whilst this system is practical for species associated with *Drosophila* spp., it is currently unclear to what extent host-associations represent distinct lineages when the *Wolbachia* genus is considered as a whole. More recently, a system for naming and typing *Wolbachia* strains on the gene *wsp* (*Wolbachia* surface protein) has been proposed (Zhou *et al.*, 1998). In this system, individual *Wolbachia* strains are given a unique name if they have a *wsp* gene that differs from other sequenced *wsp* genes by > 2.5 %. These names follow the already accepted abbreviation style *w*Host (Rousset and de Stordeur, 1994). The problem with this approach is that the *wsp* gene sequence does not distinguish between important phenotypic differences. For example, *w*No and *w*Ma, strains that naturally infect different host *Drosophila* species, have identical *wsp* sequences but different phenotypes when transfected into *D. simulans* males (Heath, 2000). It is also arbitrary in the sense that a single threshold of sequence divergence is used without any detailed understanding of the underlying ecology or selective landscape.

### 3.1.3 Phenotypic effects of *Wolbachia*

Within arthropods, *Wolbachia* are frequently known as reproductive parasites, that is, they manipulate host reproduction for their own benefit, but at the expense of host

fitness. Reproductive parasites typically skew the sex ratio of the host offspring towards infected females to increase their own transmission through the ovum. Dobson *et al.* reported that in addition to reproductive tissues, *Wolbachia* can also infect the brain, muscles, midgut, salivary glands, Malpighian tubules, fat body, wings and haemolymph (Dobson *et al.*, 2002).

In arthropods, this maternally inherited bacterium has developed a bewildering repertoire of strategies with which to manipulate host reproduction, which are discussed in more detail below. Briefly, they can affect the fertility of hosts by i) controlling compatibility between infected males and uninfected females (cytoplasmic incompatibility), ii) alter sex ratios by inducing asexual production of female offspring (parthenogenesis) or iii) by killing males (male killing) and even iv) turn males into females (feminization). However, not all symbioses are weighted in favour of the bacteria. In filarial nematodes the *Wolbachia* are strict mutualists, essential for larval and embryo development and the survival of adult worms (Taylor, 2002), and in a species of parasitic wasp which feeds on *Drosophila*, egg production is dependent on the symbiosis (Dedeine *et al.*, 2001).

Cytoplasmic incompatibility (CI) is the most common effect that *Wolbachia* can have on host reproduction. First identified in the 1970s with *Culex pipiens* (Charlat *et al.*, 2003; Yen and Barr, 1971), CI can be either unidirectional, in which embryonic lethality arises in mating between infected males and uninfected females, or bidirectional, where incompatibility arises if two partners are infected with different *Wolbachia* strains (Taylor, 2004). CI has been described in numerous insect orders including Coleoptera, Diptera, Homoptera, Hymenoptera, Orthoptera and Lepidoptera (Stouthamer *et al.*, 1999). In addition, CI has also been observed outside the Class Insecta, in Arachnida (Breeuwer and Jacobs, 1996) and Crustacea (Stouthamer *et al.*, 1999).

Parthenogenesis refers to the growth and development of an embryo or seed without fertilization by a male. With parthenogenesis-inducing *Wolbachia* (PI) the first mitotic division is abandoned within infected eggs leading to a diploid nucleus in an unfertilised egg. PI has mainly been observed within the Hymenoptera, with most research focusing on a wide range of parasitic wasps (Stouthamer *et al.*, 1993b;

Werren *et al.*, 1995b). It is within the Hymenoptera that arrhenotoky is observed, in which males arise from haploid eggs and females from diploid eggs. Therefore, the production of females from both unfertilised and fertilised eggs is observed in PI. Consequently, infected females produce twice as many daughters as uninfected females, again facilitating the spread of *Wolbachia* (Charlat *et al.*, 2003; Stouthamer *et al.*, 1999).

Male killing is also a common phenotype found and occurs early in male embryogenesis, although the mechanism remains unknown. Described in a number of Lepidoptera species in addition to *Drosophila*, *Wolbachia* male-killing provides females with a number of benefits to the surviving female progeny, including obtaining nutrients by consuming their dead brothers and by reduced competition (Charlat *et al.*, 2003; Hurst and Jiggins, 2000).

Finally, feminization, the ability of *Wolbachia* to change genetic males into females appears to be restricted to isopod crustaceans (Bouchon *et al.*, 1998), being most studied in woodlice, and also one Lepidopteron insect, *Ostrinia furnacalis* (Kageyama *et al.*, 2002). In infected populations *Wolbachia* controls sex determination, as all infected females are genetically male. The exact mechanism of feminization is unknown, although it is believed *Wolbachia* may have an effect on the androgenic hormones that control male differentiation (Taylor, 2004).

The wide range of strategies employed by *Wolbachia* present a commensurate range of fitness costs to the hosts. One putative "species" which appears to behave less like a reproductive parasite and more like a symbiont is that of *W. popcorn*. As *Wolbachia* is vertically transmitted, its survival is dependent upon host propagation, and therefore there should be selective pressure to minimise the harmful effects on the host. The identification of *W. popcorn* was therefore surprising, as this species appears to be highly virulent. Min and Benzer (1997) reported that *W. popcorn* proliferates in the brain, muscles, retina and oocytes of adult *Drosophila melanogaster*. The bacterium counteracts the fitness cost to the host by increasing host fecundity over the short-term, thus facilitating bacterial transmission, but the hosts show a significantly decreased life-span. Infected flies suffered a shortened lifespan, due to an increase in fecundity. There are two possible mechanisms by

which this increased virulence might evolve, either genes involved in the control of replication have been lost, leading to more rapid spread, or new genes have been acquired (possibly by horizontal transfer) which have increased virulence (Werren, 1997b). Either scenario makes this species an interesting and useful bacterium for studying the replication of intracellular bacteria and the mechanisms and evolution of virulence (Werren, 1997b).

### 3.1.4 Distribution of *Wolbachia*

*Wolbachia* is known to infect a wide range of hosts, and the proportion of infected individuals within a host population can be very high. Recent surveys have found *Wolbachia* infects 16.9 % – 76 % of terrestrial invertebrate species (Bouchon *et al.*, 1998; Jeyaprakash and Hoy, 2000; Werren *et al.*, 1995a). Interestingly, despite the fact that *Wolbachia* are known to infect terrestrial crustaceans, no marine *Wolbachia* have been reported to date. The large range of this estimate reflects differences in the PCR methodology used to detect *Wolbachia*, but there is a common consensus that the true figure falls towards the higher extreme. Arthropods are among the most "successful" of the animal kingdom in terms of species richness as well as abundance; their biomass far exceeds that of mammals. There are an estimated total of five million insect species on the planet (May, 2002) approximately a million of which have been formally named, and *Wolbachia* is estimated to infect 0.8M – 3.8M of these.

In the absence of a formal species nomenclature, researchers currently refer to different *Wolbachia* lineages as "supergroups". Currently, there are 8 taxonomic supergroups within the *Wolbachia* genus (Bordenstein and Rosengaus, 2005; Casiraghi *et al.*, 2005; Czarnetzki and Tebbe, 2004; Lo *et al.*, 2002; Werren *et al.*, 1995a; Rowley *et al.*, 2004), although with the number of known infected species gradually increasing, this will undoubtedly change (Table 1).

| Supergroup | Content |
|---|---|
| A | Various arthropods |
| B | Various arthropods |
| C | Filarial nematodes |
| D | Filarial nematodes |
| E | Springtails |
| F | Termites, Weevils, True bugs, Filarial nematodes |
| G | Australian spiders, Woodlouse spider, Fig wasp |
| H | Termite |

**Table 1** *Wolbachia* supergroups A – H and the host species in which they are currently found.

The supergroups are constructed using the bacterial cell-cycle gene *ftsZ* (Werren *et al.*, 1995a). This locus was considered a powerful phylogenetic marker for *Wolbachia*, as it contains both conserved regions which facilitate primer design, and more divergent regions providing increased resolution. Furthermore, 16S rRNA is also commonly used as a phylogenetic marker in *Wolbachia* due to a low degree of sequence divergence. Comparisons can be drawn between the two gene loci in order to determine the extent of phylogenetic congruence, and thus indirectly horizontal transfer. The very diverse *wsp* (*Wolbachia* surface protein) gene has also been used as a phylogenetic marker, but this gene is more routinely used as a diagnostic marker to identify cases of infection. Each of these genes will be discussed in more detail later in this chapter. Casiraghi *et al.*, (2005) recently described the phylogeny of *Wolbachia* using two further protein coding loci in addition to *ftsZ*. The housekeeping genes *gltA* and *groEL* were sequenced and their phylogenies compared to that of *ftsZ*. The study found that each of the single gene trees resolved the *Wolbachia* phylogeny well. *gltA* and *ftsZ* were both able to clearly distinguish between supergroups A – F, however *groEL* was unable to resolve supergroup E. In addition, Casiraghi concatenated the sequence data for the 3 loci, and this

concatenated phylogeny provided robust delineation between the supergroups. The majority of studies have so far focused on supergroups A & B, in which high levels of recombination have been observed between these groups. To date, no horizontal transfer has been observed between *Wolbachia* supergroups C & D, and with the remaining supergroups (Bandi *et al.*, 1998; Lo *et al.*, 2002; Panaram and Marshall, 2006).

### 3.1.5 *Wolbachia* genomics

Two *Wolbachia* genomes have been sequenced to date, *Wolbachia* endosymbiont of *Drosophila melanogaster*, *w*Mel, (Wu *et al.*, 2004) and *Wolbachia* endosymbiont strain TRS of *Brugia malayi*, *w*Bm, (Foster *et al.*, 2005), with a third in progress; *Wolbachia pipientis* endosymbiont of *Culex quinquefasciatus.*

*Wolbachia* has an unusual genomic composition in comparison to other bacteria. Firstly, like most endosymbionts it has a small, singular chromosome, indicating it has undergone extensive genome reduction throughout its evolution (Moran, 2003).

Secondly, *Wolbachia* does not exhibit the typical minimum/maximum GC skew patterns associated with other prokaryotes, which is most likely a result of intragenomic rearrangements associated with repetitive DNA (Wu *et al.*, 2004). GC skew refers to biases in base composition within the leading and lagging strand of replication. Typically, the leading strand is enriched for G (over C) and T (over A), and this is thought to result from a preponderance of C→T mutations on the leading strand. Because this bias is associated with replication, GC skew is useful for identifying the origin and terminus of replication in prokaryotic genomes. However, because this skew is not present in *Wolbachia*, the origin of replication was identified by the presence of the *dnaA* gene which is responsible for replication initiation and thought to be universally present at the replication origin.

Thirdly, both *w*Mel and *w*Bm contain large amounts of repetitive and mobile DNA, which is unique for an intracellular species as most obligate intracellular species such as *Buchnera, Chlamydia, Rickettsia and Wigglesworthia* all have low levels of repetitive and mobile DNA (Wu *et al.*, 2004). Repetitive DNA is commonplace in

eukaryotes, but relatively rare in prokaryotes. The length, sequence and position of these sequences in the genome are highly variable, with 14.2 % of the *w*Mel genome represented by repeats of 200 bp or more. This variability in sequence and copy number is different between *Wolbachia* strains making it a potentially useful marker for strain discrimination (Wu *et al.*, 2004). The *w*Bm genome contains considerably less repetitive DNA than in the *w*Mel genome, which may be a result of stronger selection pressure to remove parasitic mobile elements in *w*Bm as there is little evidence of exposure to foreign DNA (Foster *et al.*, 2005). Comparison of the repetitive elements between these two genomes suggests the invasion of mobile genetic elements occurred after the divergence of the two *Wolbachia* along the *w*Mel branch, or that the majority were successfully eliminated within the *w*Bm branch (Foster *et al.*, 2005). In addition, extensive genome shuffling has occurred within *w*Bm and *w*Mel, eliminating gene order (synteny) between the two genomes. This is most likely due to the large number of repetitive DNA and active recombination events, which have been identified in large numbers in the wMel genome (Foster *et al.*, 2005; Wu *et al.*, 2004).

Finally, the *w*Mel genome contains unusually large numbers of genes that encode proteins containing ankyrin repeat domains (Wu *et al.*, 2004), one of the most common protein-protein interaction motifs in nature. Ankyrin repeats are tandemly repeated motifs of about 33 amino acids (Iturbe-Ormaetxe *et al.*, 2005; Mosavi *et al.*, 2004). While these domains are common in both eukaryotic and viral proteins, they are relatively rare in bacteria, usually only present in only a few copies per species. The few known examples from prokaryotes and viruses may be the result of horizontal gene transfers (Iturbe-Ormaetxe *et al.*, 2005; Wu *et al.*, 2004). *w*Bm has only five ankyrin repeat genes, whereas *w*Mel has twenty-three repeat genes which is particularly unusual due to its small genome size. As yet it is still unclear as to why *w*Bm and *w*Mel have different numbers of ankyrin repeat genes. In eukaryotes ankyrin repeat domains mediate protein interactions in a wide spectrum of proteins, including cytoskeletal and membrane proteins, transcriptional and developmental regulators, toxins and cyclin-dependent kinase inhibitors. The precise function of the repeats in bacteria is still largely unknown, however they may play a role in host cell-cycle regulation (Caturegli *et al.*, 2000), cytoplasmic incompatibility (Tram and

Sullivan, 2002) and interaction with host cytoskeleton (Hryniewicz-Jankowska *et al.*, 2002).

### 3.1.6 Phylogeny of *Wolbachia*

Phylogenies of *Wolbachia* have been constructed using a number of loci. Initial studies generally used 16S rRNA as a molecular marker (Rousset *et al.*, 1992b). On the basis of 16S rRNA sequence comparisons, *Wolbachia* clusters with members of the family *Anaplasmataceae*; *Anaplasma, Ehrlichia, Neorickettsia* and *Wolbachia* (Dumler *et al.*, 2001). The close relationship between *Wolbachia* and these other species is curious, as *Wolbachia* strains are not known to be associated with vertebrate disease, something which is ubiquitous to the other genera of this family (Anderson and Karr, 2001). Despite the variety of effects that *Wolbachia* exerts on its diverse hosts, there is less than 3 % difference in the 16S rRNA from several *Wolbachia* strains (O'Neill *et al.*, 1992a; Stouthamer and Werren, 1993a), indicating the locus is too conserved to give much phylogenetic resolution within *Wolbachia* spp. A lack of congruence between host association and phylogenetic relationship based on the 16S rRNA gene has been inferred as evidence for frequent horizontal transmission (O'Neill *et al.*, 1992a). However, the high degree of sequence conservation, and slow rate of evolution, means that 16S rRNA provides limited evidence regarding population structure, patterns of diversity and intertaxon transmission (Werren, 1997a; Zhou *et al.*, 1998) and should only be used to discriminate between broad groupings and not between individual *Wolbachia* species.

The *ftsZ* cell cycle gene, which evolves more rapidly than 16S rRNA has proved a useful marker for studying fine scale relationships between *Wolbachia* strains found in diverse hosts (Schilthuizen and Stouthamer, 1997; Werren *et al.*, 1995a; Werren *et al.*, 1995b). The *ftsZ* locus has confirmed the existence of 8 taxonomic "supergroups" within the *Wolbachia* genus. There is broad concordance between these supergroups and the bacterial host, most arthropod-derived strains are classified into two major clades, A and B (Werren *et al.*, 1995b). Similarly, supergroups C and D are composed of strains associated with filarial nematodes (Table 1).

Although the phylogenies based on 16S rRNA and *ftsZ* sequences show broadly consistent groupings, neither gene has provided sufficient information to adequately resolve the relationships between individual *Wolbachia* strains that display different reproductive phenotypes but infect the same host or to examine in detail the concordance between bacterial and host phylogenies (Jiggins *et al.*, 2001; Stouthamer *et al.*, 1999; Werren, 1997a). Recent reports have suggested that the phylogeny of *Wolbachia* does not tightly correspond to host species, beyond the broad groupings of arthropods and nematodes described above, which implies that *Wolbachia* must utilise a combination of vertical and horizontal transmission (O'Neill *et al.*, 1992a; Rousset *et al.*, 1992b; Werren *et al.*, 1995a; Werren *et al.*, 1995b; Zhou *et al.*, 1998). The possibility of horizontal transmission remains contentious, however, as *Wolbachia* has never been observed to be free living, and it is unclear how this might occur.

In an attempt to improve resolution, several other phylogenies have been constructed, including 23S rRNA (Rousset *et al.*, 1992a), spacer 2 region (van Meer *et al.*, 1999a) and *groEL* (Anderson and Karr, 2001; Casiraghi *et al.*, 2005). The *wsp* gene, which codes for an outer membrane protein of *Wolbachia* (Braig *et al.*, 1998), also exhibits higher variation than *ftsZ,* shows more divergence between and within the *Wolbachia* supergroups than other loci and has the advantage of an extensive sequence data base available (van Meer *et al.*, 1999b; Zhou *et al.*, 1998). However, localisation of the *wsp* protein at the interface between two cellular environments and the presence of regions under strong positive selection results in recombination events being common place within the gene (Baldo *et al.*, 2005b). Baldo *et al.*, (2005b) found evidence of extensive recombination in *Wolbachia* surface protein, which had resulted in shuffling of hypervariable region motifs within the protein. Also, *wsp* has undergone both intragenic and intergenic recombination which will have the effect of producing artifactual phylogenies (Baldo *et al.*, 2005b). For this reason, *Wolbachia* phylogeny cannot be inferred solely on the basis of one gene which is frequently recombining. This underscores the necessity of using multiple coding loci in reconstructing intragenus phylogenies (Casiraghi *et al.*, 2005).

### 3.1.7 Horizontal transmission

The *ftsZ* phylogeny clearly shows horizontal transmission within both A and B *Wolbachia* supergroups, and indicates that particular strains, such as Adm have undergone extensive horizontal transfer (Werren *et al.*, 1995b). *Wolbachia* of the *Adm* strain taken from Hymenoptera, Lepidoptera and Coleoptera show virtually identical *ftsZ* sequences despite the fact that these hosts diverged from each other around 200 million years ago (Heath, 2000), indicating that extensive horizontal transfer has occurred in this group (Werren *et al.*, 1995b). A possible indication of exchange between parasitoid and host is given by an almost identical *ftsZ* sequence in *Drosophila melanogaster* and the Drosophilid larval parasitoid *Asobara tabida* (Werren *et al.*, 1995b). Lower levels of horizontal transfer are found among B supergroup strains, as evidenced by greater congruence with host phylogenies (Werren *et al.*, 1995b). The *ftsZ* phylogeny reveals very closely related B supergroup *Wolbachia* in *Nasonia girulti* wasps and their *Protocalliphora* (Diptera) hosts, which may indicate parasitoid mediated horizontal transfer (Werren *et al.*, 1995b). The phylogenies of nematodes and their *Wolbachia* (supergroups C & D) demonstrate congruence, which is in keeping with the strict vertical transmission of the microbe and the obligate association between the partners.

Horizontal transfer of *Wolbachia* has been observed between *Leptopilina boulardi*, an endoparasitoid of Drosophilid larvae, and *D. simulans* (Riverside) which had been cured of *Wolbachia* by tetracycline treatment (DSRT). This was the first natural horizontal transfer route for *Wolbachia* between phylogenetically distinct insect species (Heath *et al.*, 1999). The hypothesis of parasitoid mediated horizontal transfer is however not supported by all phylogenies. For example, sympatric sibling species such as *D. melanogaster* and *D. simulans*, which share common parasitoids, do not share the same *Wolbachia* infections (Heath, 2000).

Multiple infections of *Wolbachia* are well documented and are known to exist in several insect taxa including parasitic wasps, *Drosophila*, leafcutter ants and bruchid beetles (Hiroki *et al.*, 2004). Double infections are abundant, but naturally occurring triple or more infection is rare (Jamnongluk *et al.*, 2002) although the experimental establishment of triple infection has been reported (Hiroki *et al.*, 2004). Theoretical considerations predict that multiple infections should result in an increase in

virulence due to competition between strains, and thus the vertical transmission of symbionts tends to limit multiple infections. Thus, with strict vertical transmission multiple infections within host individuals should be transient and be selectively unfavoured over evolutionary time. However, the possibility of horizontal transmission might result in a higher degree of mixed infections (as multiple strains can infect a single host), and a higher level of virulence (due to competition between strains and a lower trade-off for transmission opportunities) (Mouton *et al.*, 2004). A high degree of mixed infection will also provide greater opportunities for gene transfer between divergent lineages (Werren, 1997a).

As mentioned above, the mechanism of horizontal transmission between hosts remains unclear, but one possibility is that hosts may be infected orally by ingestion of infected prey. Spiders are particularly attractive for studying this possibility given their interaction with prey from across arthropod lineages. Under a model of predator:prey transmission, spiders would be expected to harbour a diversity of *Wolbachia* strains reflecting those found in prey species (Rowley *et al.*, 2004).

## 3.2 METHODS

### 3.2.1 Sampling

This study is largely based on a wide range of host species sampled across four continents between 1996 and 2004 by Dr. Robert Butcher (then of University of Bath; Fig. 3). The host species encompassed five Orders from the Class Insecta (Hymenoptera, Coleoptera, Lepidoptera, Diptera, Isoptera), in addition to one Order from the Class Arachnida (Acari). A number of hosts from laboratory stocks across Europe were included, including *Drosophila* and Hymenoptera (Appendix A) also collected by Dr. Robert Butcher.

### 3.2.2 Isolation of samples

Samples were recovered as described previously in Chapter 2, Section *2.2 Preparation and Storage of Cell and DNA stocks.*

### 3.2.3 Isolate nomenclature

Each individual *Wolbachia* isolate was given a unique two-part code. The first part refers to the host species, whereas the second is a strain identifier. For example, isolate PW07, is strain 7 isolated from a parasitic wasp. Other host abbreviations are listed in Table 2 in addition to host species and their respective common names.

| Common name | ID prefix | Species name | Order: Family |
|---|---|---|---|
| Not determined | A | Not determined | Hymenoptera: Formicidae |
| Cornflour beetle | B | *Tribolium confusum* | Coleoptera: Tenebrionidae |
| Redflour beetle | B | *Tribolium castaneum* | Coleoptera: Tenebrionidae |
| Not determined | BEE | Not determined | Hymenoptera: Apidae |
| Not determined | BF | Not determined | Lepidoptera: Hesperiidae |
| Cabbage white butterfly | BF | *Pieris rapae* | Lepidoptera: Pieridae |
| Fruit fly | F | *Drosophila simulans*<br>*Drosophila subobscura*<br>*Drosophila simulans/ melanogaster* hybrid | Diptera: Drosophilidae |
| Not determined | HF | Not determined | Diptera: Syrphidae |
| Diamondback moth | M | *Plutella xylostella* | Lepidoptera: Plutellidae |
| Nettle tap moth | M | *Anthophila fabriciana* | Lepidoptera: Choreutidae |
| Dark-sword grass moth | M | *Agrotis ipsilon* | Lepidoptera: Noctuidae |
| Parasitic wasp to the European corn borer | PW | *Macrocentrus grandii* | Hymenoptera: Braconidae |
| Parasitic wasp to the common grain moth | PW | *Bracon herbetor* | Hymenoptera: Ichneumonidae |
| Parasitic wasp to the beet webworm | PW | *Phytodietus niger* | Hymenoptera: Ichneumonidae |
| Parasitic wasp to the fruit fly | PW | *Asobara tabida* | Hymenoptera: Braconidae |
| Parasitic wasp to the budworm | PW | *Campoletis sonorensis* | Hymenoptera: Ichneumonidae |
| Parasitic wasp to the dark beetle | PW | *Roptrocerus xylophagorum* | Hymenoptera: Pteromalidae |
| n/a | SEF | *Olescampe mascellator* | Diptera: Diopsidae |
| n/a | SEF | *Cyrtodiopsis dalmanni* | Diptera: Diopsidae |
| n/a | SEF | *Cyrtodiopsis whitei* | Diptera: Diopsidae |
| n/a | SEF | Not determined | Diptera: Diopsidae |
| Pine saw-fly | S | *Diprion similis* | Hymenoptera: Diprionidae |
| Spider mite | SM | *Bryobia sarothamni* | Acari: Tetranychidae |
| Not determined | T | Not determined | Isoptera |
| Paper wasp | W | *Polistes fuscatus* | Hymenoptera: Vespidae |

**Table 2** Host species and identification prefix used in this MLSA study; A=ant; B=beetle; BEE=bee; BF=butterfly; F=fly; HF=hoverfly; M=moth; PW=parasitic wasp; SEF=stalk-eyed fly; S=saw-fly; SM= spider mite; T=termite; W=wasp.

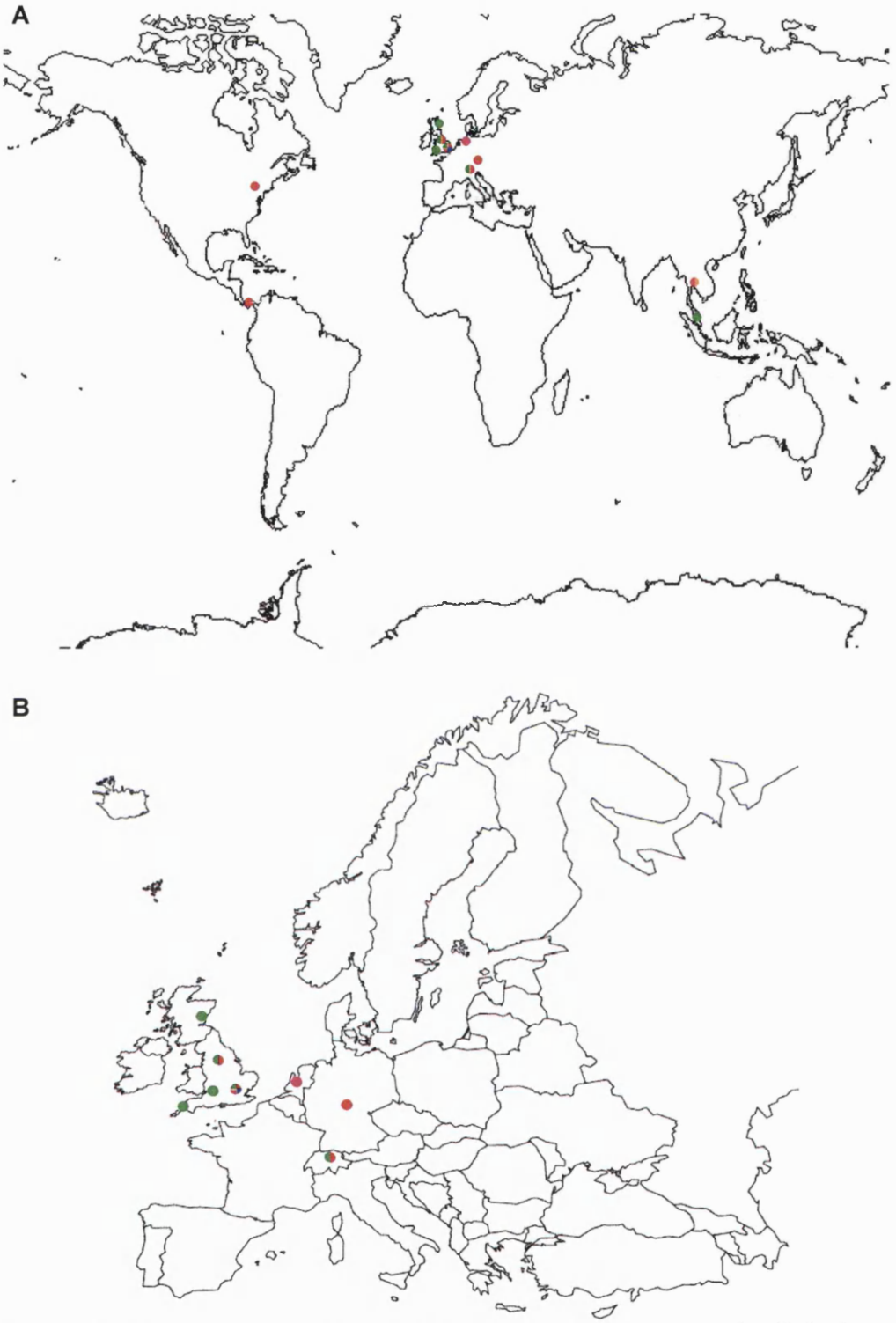



**Figure 3** Sampling locations across the globe (a) and a more detailed view of Europe (b). Each location is represented by a circle, the colour of which represents the Order of the host species; red–Hymenoptera; blue–Coleoptera; green-Lepidoptera; pink–Diptera; purple–Acarina; orange–Isoptera. Sites at which multiple Orders were collected are represented by circles with each representative colour.

### 3.2.4 Choice of genes, primer design & PCR conditions

A total of 21 primer pairs were tested, of which 5 were used, as they produced the most reliable PCR and sequencing results, in addition to *wsp* & *ftsZ* (Table 3). The MLSA loci selected are also evenly distributed around the *Wolbachia pipientis* *w*Mel genome and in the majority of cases located beside other conserved loci (Fig. 4). The loci chosen therefore fulfil the criteria suggested for MLSA genes (Cooper and Feil, 2004). PCR and sequencing was carried out as previously described (Chapter 2 Section 2.4 *Methods for Polymerase Chain Reaction and Sequencing*).

| Loci | Gene function | Primer design method |
|---|---|---|
| *atpA* | Encodes the alpha subunit of ATP synthase. | By reference to genome sequence. |
| *gltA* | Catalyses the conversion of acetyl CoA with oxaloacetate to form citrate & coenzyme A. | By reference to GenBank. |
| *groEL* | Assists with protein folding in an ATP-dependent manner. | By reference to GenBank. |
| *tpiA* | Encodes a glycolytic enzyme that catalyzes the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. | By reference to genome sequence. |
| *trmD* | Encodes the enzyme tRNA (guanosine-1)-methyltransferase. | By reference to genome sequence. |
| *wsp* | *Wolbachia* surface protein. | (Zhou *et al.*, 1998) & modified by Dr. Robert Butcher. |
| *ftsZ* | Essential cell division gene involved in the regulation of cell division. | Donated by Dr. Robert Butcher. |

**Table 3** Gene function & primer design method for each locus.

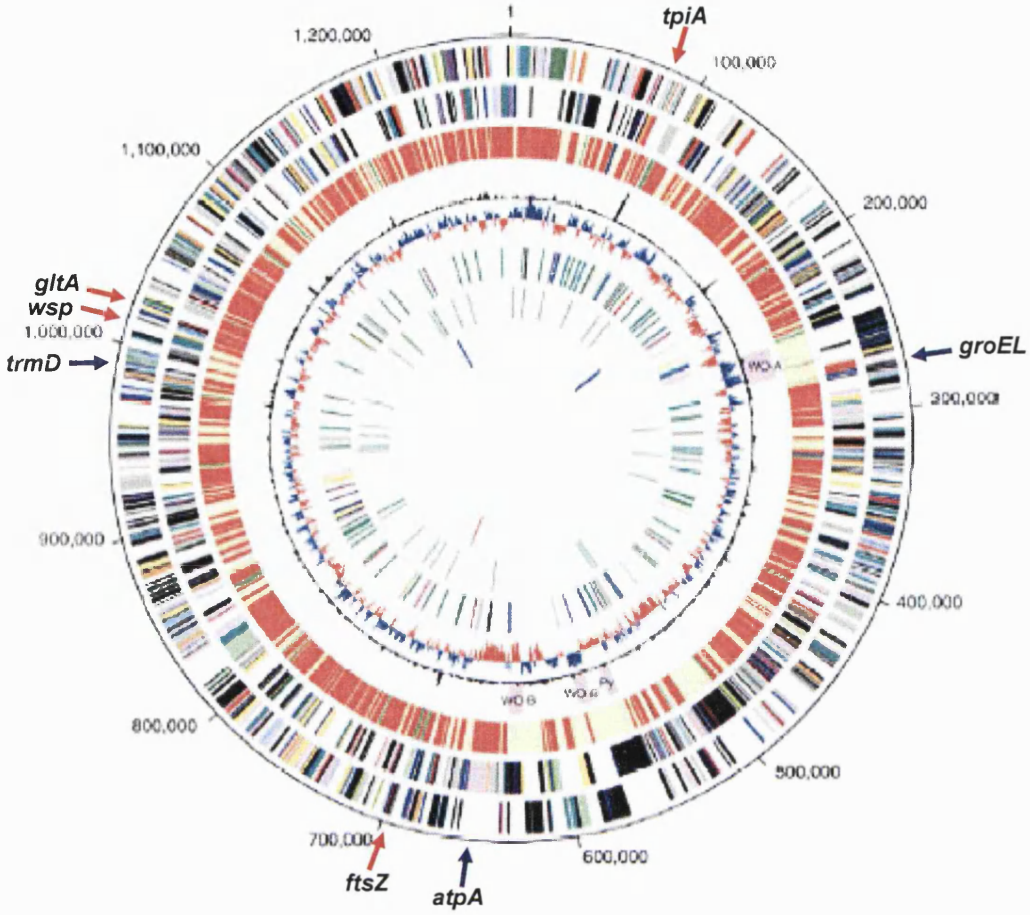



**Figure 4** *Wolbachia pipientis w*Mel genome labelled with *Wolbachia* MLSA genes and traditional *Wolbachia* spp. phylogenetic markers. Genes located on the forward and reverse strand are coloured red and blue respectively. Modified figure (Wu *et al.*, 2004).

PCR conditions for each loci varied by annealing temperature and extension time (Table 4). All primer sequences can be found in Appendix A.

| Loci | Annealing | | Extension | |
|---|---|---|---|---|
| | Temperature | Time | Temperature | Time |
| *atpA* | 58°C | 1 min | 72°C | 1 min |
| *gltA* | 52°C | 1 min | 72°C | 1.5 min |
| *groEL* | 49°C | 1 min | 72°C | 1.5 min |
| *tpiA* | 52°C | 1 min | 72°C | 1 min |
| *trmD* | 55°C | 1 min | 72°C | 1.5 min |
| *wsp* | 60°C | 1 min | 72°C | 1.5 min |
| *ftsZ* | 48°C | 1 min | 72°C | 1.5 min |

**Table 4** PCR conditions of each locus.

3.2.5 Nucleotide sequence analysis

-Sequence editing and alignment

As described in Chapter 2 Section 2.5 *Nucleotide sequence analysis.*

3.2.6 Phylogenetic analysis

-Distance methods

As described in Chapter 2 Section 2.6 *Phylogenetic analysis.*

-Bayesian inference of phylogeny

As described in Chapter 2 Section 2.6 *Phylogenetic analysis.*

The following parameters were used to create the Bayes block:

- number of generations = 2,000,000
- number of chains = 4
- sampling frequency = 1000
- print frequency = 100

- burnin = 20%.

### 3.2.7 Tests for recombination

-Tests of neutrality

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Sawyer's Run Test

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Population-scaled recombination rate (ρ)

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-DnaSp

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Splits decomposition

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

### 3.2.8 Other methods of analysis

-eBURST

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

-Datamonkey

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

## 3.3 RESULTS

### 3.3.1 The dataset

A total of 44 strains were isolated and sequenced at 5 housekeeping loci (Table 5); *atpA, gltA, groEL, trmD & tpiA*. In addition, the gene encoding the *Wolbachia* surface protein, *wsp*, was sequenced in 26 strains. For details of eukaryotic host and geographical location, please see Appendix A.

| Order | Species name | No. of species |
|---|---|---|
| Acarina | *Bryobia sarothamni* | 1 |
| Coleoptera | *Tribolium castaneum* | 1 |
| Coleoptera | *Tribolium confusum* | 2 |
| Diptera | *Cyrtodiopsis dalmanni* | 1 |
| Diptera | *Cyrtodiopsis whitei* | 1 |
| Diptera | *Drosophila simulans* | 3 |
| Diptera | *Drosophila simulans/melanogaster hybrid* | 1 |
| Diptera | *Drosophila subobscura* | 1 |
| Diptera | *Olescampe mascellator* | 1 |
| Diptera | Unknown | 2 |
| Hymenoptera | *Asobara tabidia* | 1 |
| Hymenoptera | *Bracon hebetor* | 1 |
| Hymenoptera | *Campoletis sonorensis* | 1 |
| Hymenoptera | *Diprion similis* | 1 |
| Hymenoptera | *Macrocentrus grandii* | 1 |
| Hymenoptera | *Phytodietus niger* | 1 |
| Hymenoptera | *Polistes fuscatus* | 1 |
| Hymenoptera | *Roptrocerus xylophagorum* | 1 |
| Hymenoptera | Unknown | 4 |
| Isoptera | Unknown | 1 |
| Lepidoptera | *Agrotis ipsilon* | 1 |
| Lepidoptera | *Anthofilia fabriciana* | 5 |
| Lepidoptera | *Helicoverpa armigera* | 1 |
| Lepidoptera | *Pieris rapae* | 1 |
| Lepidoptera | *Plutella xylostella* | 7 |
| Lepidoptera | Unknown | 1 |

**Table 5** Host species from which *Wolbachia* spp. were isolated (n=44).

3.3.2 Sequence parameters

The alleles defined for the study were between 366 bp (*gltA*) and 615 bp (*wsp*) in length, and between 11 (*atpA*) and 27 (*groEL*) alleles were present per locus (Table 6). In order to gauge the selective pressure of the protein-coding genes, the dS/dN ratio was calculated as described in Chapter 1. *atpA* is the most evolutionary conserved protein-coding gene in that it exhibits both the highest dS/dN ratio (12.10), the lowest average pairwise divergence ($\pi = 0.037$), the lowest percentage of variable sites (11.4 %) and the lowest number of alleles. It is therefore under the greatest level of purifying selection of the five housekeeping genes. *gltA* is the second most conserved locus with a dS/dN ratio of 9.17. Although *groEL* has a slightly higher dS/dN ratio (10.77) to that of *gltA,* suggesting it is under greater purifying selection, it has a higher level of $\pi$ (0.075 & 0.051, respectively), a greater percentage of variable sites (23.6 % & 21.6 %, respectively) and a higher number of alleles for the gene (27 & 19, respectively). With a low dS/dN ratio (4.45) and moderate level of $\pi$ (0.055), *tpiA* exhibits lower levels of purifying selection. Of the six loci analysed in this study, the second highest percentage of variable sites is observed in *tpiA*, with a quarter of the gene segment sequenced being variable (25.9 %). *trmD* has the second lowest dS/dN ratio of the loci (2.95) and exhibits low selective pressure. Approximately one quarter of this gene segment sequenced is variable, in addition to a $\pi$ value of 0.061. The lowest dS/dN ratio is observed with *wsp* (1.60), indicating it has the highest proportion of non-synonymous changes and the lowest level of purifying selection of the six loci. The *wsp* locus contains four hypervariable regions interspersed by conserved regions, spanning 40.8 % of the gene (Fig. 5). The 615 bp fragment sequenced contains all of the hypervariable regions and with the highest percentage of variable sites (38.5 %) and $\pi$ (0.146), *wsp* is the most divergent of the genes. However, only 26 *Wolbachia* strains were sequenced at *wsp*, therefore it is possible these values do not reflect the true level of diversity within the gene.

| Loci | Fragment size (bp) | No. alleles | % Variable sites | $d_S/d_N$ | $\pi$ |
|---|---|---|---|---|---|
| *atpA* | 402 | 11 | 11.4 | 12.10 | 0.037 |
| *gltA* | 366 | 19 | 21.6 | 9.17 | 0.051 |
| *groEL* | 492 | 27 | 23.6 | 10.77 | 0.075 |
| *tpiA* | 402 | 24 | 25.9 | 4.45 | 0.055 |
| *trmD* | 456 | 20 | 24.3 | 2.95 | 0.061 |
| *wsp* | 615 | 15 | 38.5 | 1.60 | 0.146 |

**Table 6** Genetic diversity of *Wolbachia* spp. MLSA loci (n=44) & *wsp* (n=26).

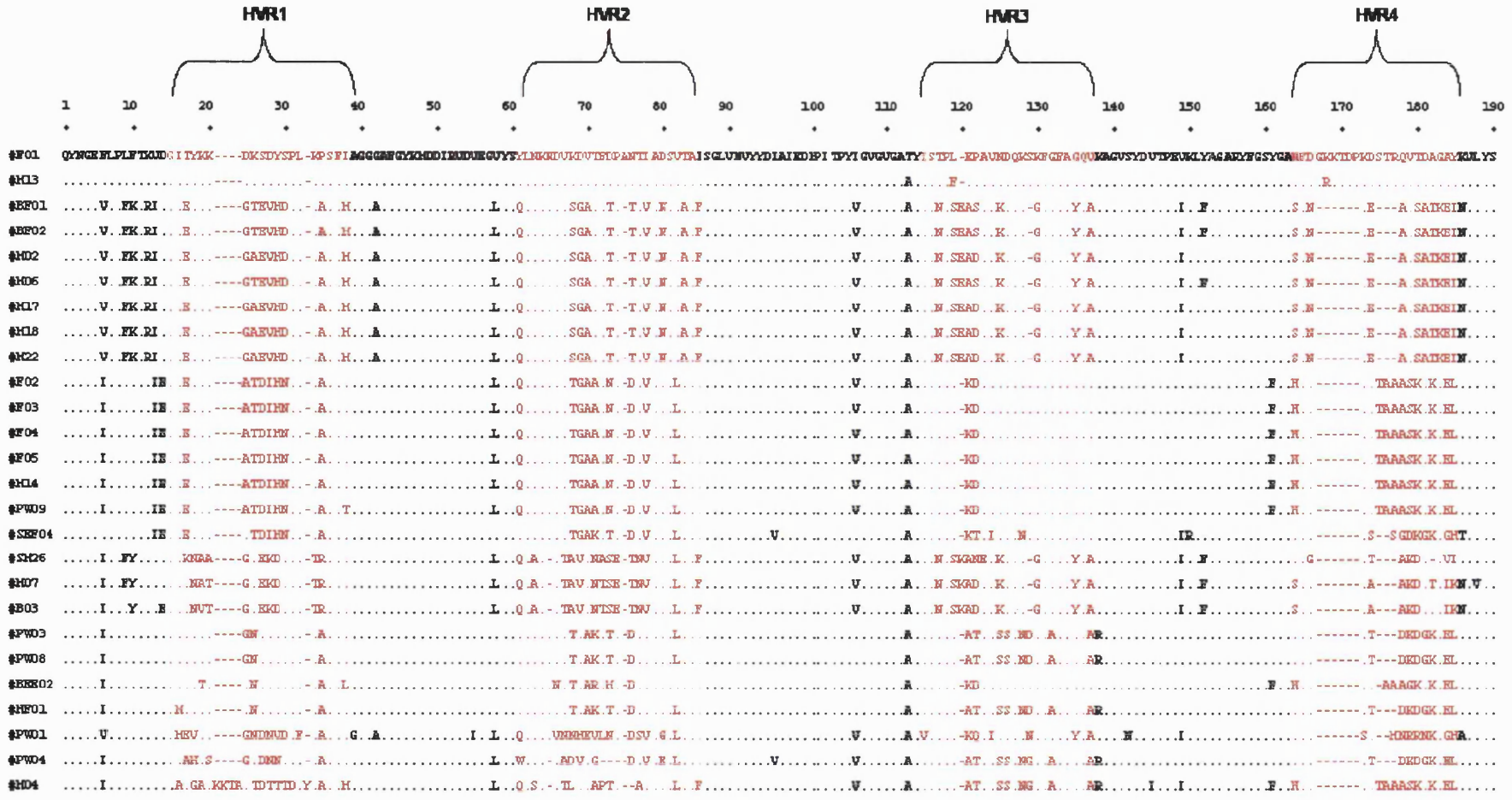



**Figure 5** Amino acid alignment of sequenced fragment of *wsp* (190 amino acids in length). Hypervariable regions (HRV) are highlighted in red according to the limits set by Baldo *et al.*, (2005b) interspersed by conserved regions.

Watterson's population mutation rate (θ) and Hudson's R parameter were calculated for each of the MLSA loci and *wsp* (Table 7). Of the five MLSA genes, *atpA* and *gltA* have the lowest population mutation rates (14.683 & 18.597, respectively) and population recombination rates (0.001 & 0.2, respectively), with *atpA* being the most evolutionary conserved gene. Both *groEL* and *trmD* exhibit similar levels of mutation rate, however the former has a higher population mutation rate (1.4) than that of *trmD* (0.001). *groEL* has been reported to be under positive selection in *Buchnera* because of its role in protein folding, but it is unclear whether this gene will be under the same selective pressures in *Wolbachia* (Wernegreen and Moran, 1999). *tpiA* has the highest population recombination rate of the housekeeping genes, and a relatively low population mutation rate (21.300). *wsp* exhibits the highest rates of recombination (20.8) and mutation (73.305) of the six loci, which is consistent with its role as an outer surface protein and with having hypervariable regions in the gene. However, these values must be carefully assessed as they may not represent the true rates due to only 26 samples available for use in this analysis.

| Loci | θ | R |
|---|---|---|
| *atpA* | 14.683 | 0.001 |
| *gltA* | 18.597 | 0.2 |
| *groEL* | 36.910 | 1.4 |
| *tpiA* | 21.300 | 2.3 |
| *trmD* | 27.617 | 0.001 |
| *wsp* | 73.305 | 20.8 |

**Table 7** Population mutation rate (θ) & recombination rate (R) of *Wolbachia* spp. MLSA loci (n=44) & *wsp* (n=26).

Tajima's D and Fu and Li's D* & F* tests were calculated to assess the level of neutrality within the six loci (Table 8). Of the six loci, only *trmD* showed evidence of selection ($0.01 > P > 0.05$) with Fu & Li's D* test, with the remaining loci failing to provide significant evidence of selection in each of the three neutrality tests ($P > 0.01$). Tajima's D test identified three positive loci (*atpA, groEL,* and *trmD*) and three negative loci (*gltA, tpiA* and *wsp*), although none of these observations were significant. Only one locus was positive for Fu and Li's F* test (*atpA*) and again all loci were not significant from zero ($P > 0.10$). Fu and Li's D* test revealed only one

positive locus, however *trmD* was the only gene to show evidence of selection (0.01 > P > 0.05), suggesting that there is perhaps an excess of rare polymorphisms, consistent with positive selection.

| Loci | Tajima's D | Fu & Li's F* | Fu & Li's D* |
|---|---|---|---|
| *atpA* | 1.15896 (P > 0.10) | 0.052795 (P > 0.10) | 0.04976 (P > 0.10) |
| *gltA* | -0.21339 (P > 0.10) | -1.30186 (P > 0.10) | -1.58998 (P > 0.10) |
| *groEL* | 1.07157 (P > 0.10) | -0.06680 (P > 0.10) | -0.70352 (P > 0.10) |
| *tpiA* | -0.48541 (P > 0.10) | -1.51080 (P > 0.10) | -1.71292 (P > 0.10) |
| *trmD* | 0.06569 (P > 0.10) | -1.46385 (P > 0.10) | -1.96805 (0.10 > P > 0.05) |
| *wsp* | -0.01800 (P > 0.10) | -0.03505 (P > 0.10) | -0.03538 (P > 0.10) |

**Table 8** Tajima's D, Fu & Li's F* and Fu & Li's D* test of *Wolbachia* spp. MLSA loci (n=44) & *wsp* (n=26).

### 3.3.3 Phylogenetic analysis

Bayesian trees were constructed for each locus as described in Methods (Fig. 6a-e) and for the concatenated sequences (Fig. 6f). The sequence analysis above revealed *atpA* to be to most uniform gene with the lowest rate of recombination. This gene was therefore used as a "standard" tree with which to compare the other genes. The *atpA* phylogeny reveals two major groups (Fig. 6a). The larger group, highlighted in red and at a basal position in Figure 6a consists of 27 isolates recovered from a wide range of hosts including an ant, beetle, bees, flies, hoverfly, moths, parasitic wasps, stalk-eyed flies, saw-fly and termite. The second group is smaller (n = 12) and consists of isolates recovered from beetles, butterflies, moths and a spider mite. This group, shown in blue in Figure 6a, is well differentiated from the larger group (posterior probability = 100%). Intermediate to these two groups lie a further 5 strains, which were recovered from a moth, parasitic wasps, stalk-eyed fly and wasp. This group is more poorly supported and is coloured in green in Figure 6a.

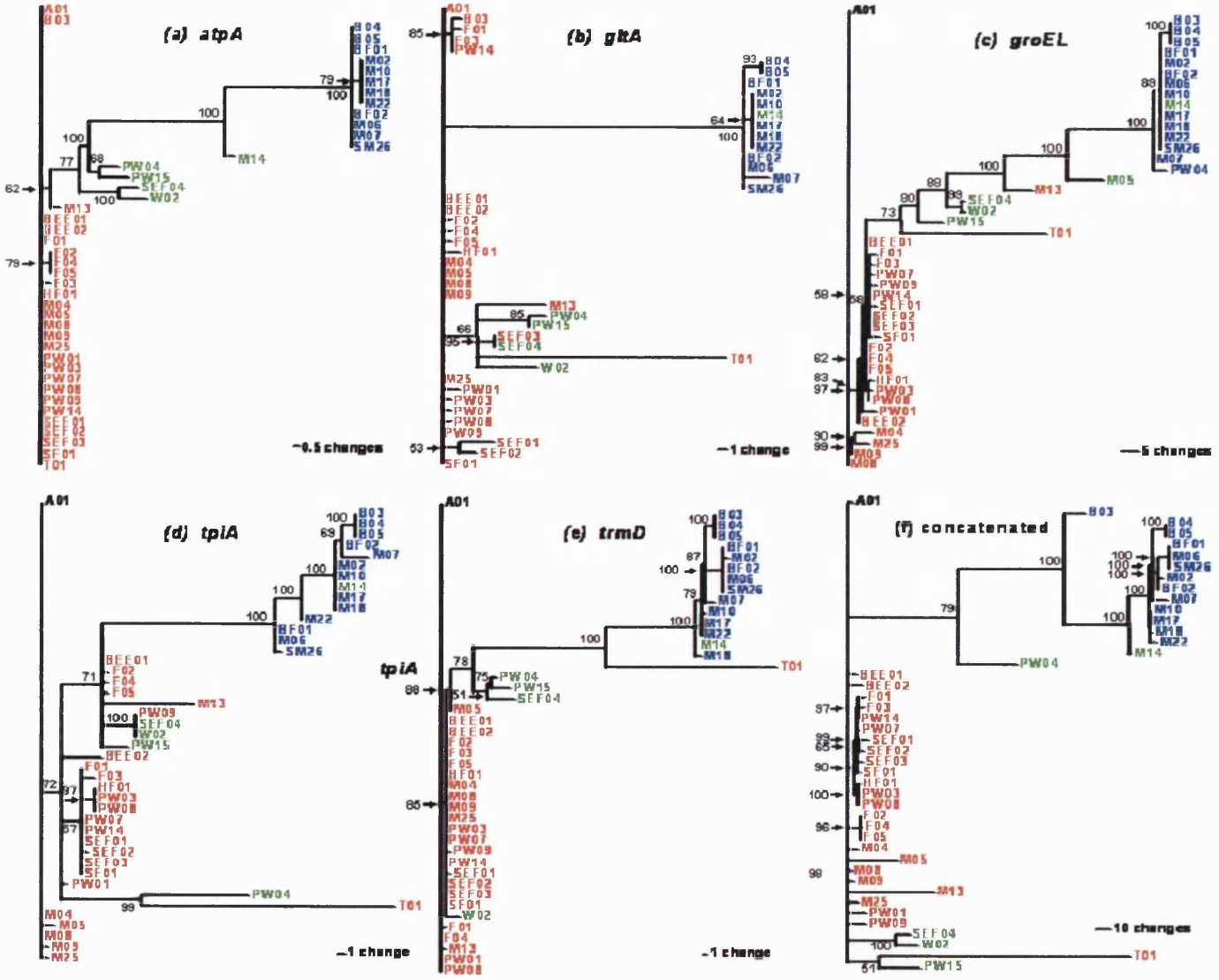



**Figure 6** Phylogenetic trees based upon the Bayesian method for each of the 5 MLSA loci (n=44), the concatenated alignment of the MLSA loci (n=44) and *wsp* (n=26). Each strain is coded according to its invertebrate host; A = ant; B = beetle; BEE = bee; BF = butterfly; F = fly; HF = hoverfly; M = moth; PW = parasitic wasp; SEF = stalk-eyed fly; SF = saw-fly; SM = spider mite; T = termite; W = wasp. Each strain is colour coded according to its position on the phylogenetic tree, i.e., red, green or blue.

In order to visually compare the topologies of the other gene trees, all strains were colour coded according to their clustering at *atpA*. The blue group was found to be consistent in all genes, except that M13 corresponded to this group at all genes except *atpA*. M14 at *atpA* appears to be a hybrid of the two main groups, consistent with recombination, however visual inspection of the polymorphic sites in this strain did not reveal significant mosaic structure (data not shown). PW04 also shows inconsistent phylogenetic position on the different gene trees, and clusters with the blue clade at *groEL*, but with the red or green (intermediate) clades at the other loci (see Figure 8 for the polymorphic sites in this strain at *groEL*). Similarly, strain B03 clusters with the red clade at *atpA* and *gltA*, but with the blue clade at *groEL*, *tpiA* and *trmD*. Other inconsistencies are restricted to topological differences between the large red clade and the intermediate and poorly supported green clade.

The concatenated tree (Fig. 6f) resolves the inconsistencies within single gene loci, and more clearly distinguishes the two main groups of isolates (red + green and blue groups). B03 and PW04 are shown in an intermediate position due to the inconsistencies described above.

### 3.3.4 Host association observed amongst isolates

The two major groups revealed by the phylogenetic analysis above both contain isolates from a wide variety of hosts. In some cases it is clear that isolates from related hosts fall in to both groups. For example, both the red and the blue groups described above contain isolates recovered from moths (Fig 6). In order to examine host association in more detail, the strains were colour coded according to host source (Fig. 7 a-g). Clustering is observed across the majority of loci, with a few exceptions. For example, three Coleoptera isolates (B03, B04, B05) appear identical at all genes with the exception of *atpA* and *gltA*, where isolate B03 has acquired a red group allele. Two Lepidopteron butterflies, BF01 and BF02, observed within the blue group, are identical at all loci with the exception of *tpiA*. A large scale recombination event has occurred across invertebrate Orders between isolate BF01 and the red group resulting in this isolate having a mosaic structure at *tpiA* (Fig. 8), thus placing it in a more basal branch of the tree than BF02.

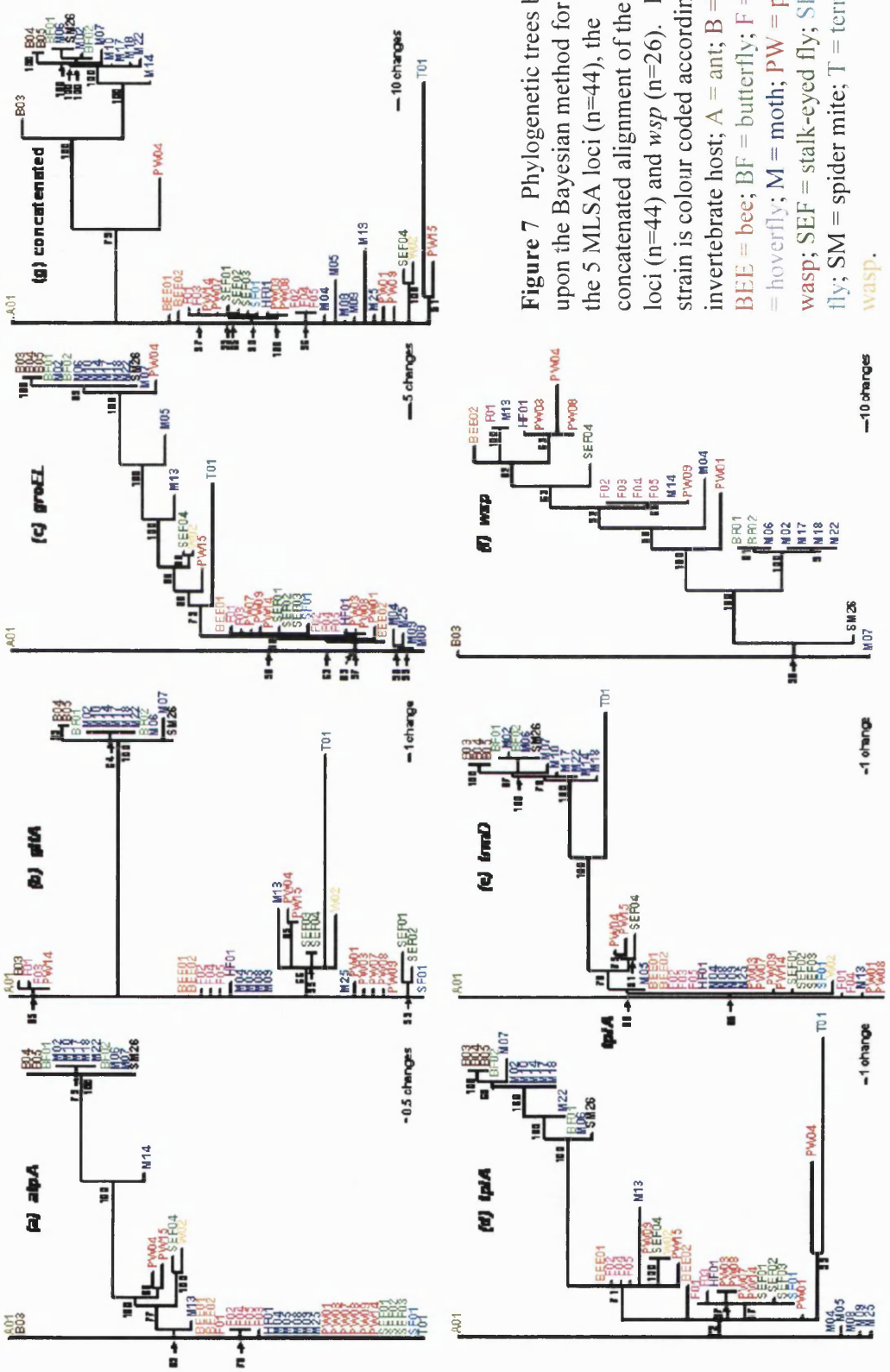



**Figure 7** Phylogenetic trees based upon the Bayesian method for each of the 5 MLSA loci (n=44), the concatenated alignment of the MLSA loci (n=44) and *wsp* (n=26). Each strain is colour coded according to its invertebrate host; A = ant; B = beetle; BEE = bee; BF = butterfly; F = fly; HF = hoverfly; M = moth; PW = parasitic wasp; SEF = stalk-eyed fly; SF = saw-fly; SM = spider mite; T = termite; W = wasp.

```
                        11111111 1222222233 33333
                111233344 5801137889 9125668912 36789
                9018105689 4621993032 7398262427 13247
#A01    AGGTCCAAGC GGTTCAAAAG CGAGGTCTTG GTCCG
#BF01   ........AT AACAGGGGGA GACAACACCA ACTAA
#BF02   GAACTTGGAT AACAGGGGGA GACAACACCA ACTAA
```

**Figure 8** Evidence of mosaicism within *tpiA* (BF01) resulting from horizontal transfer between Hymenoptera (A01) and Lepidoptera (BF02).

The two Hymenopteran Apidae isolates (BEE01, BEE02) are identical at all the MLSA loci except for *tpiA*, where recombination has taken place, however are not of a mosaic structure (data not shown). Of the four Dipteran Diopsidae isolates, SEF01, SEF02 and SEF03 are identical or very nearly identical at all the loci, with SEF04 remaining very dissimilar. However, at *gltA* SEF03 & SEF04 and SEF01 & SEF02 cluster separately.

The sample also contains multiple isolates from moths (n = 14) and parasitic wasps (n = 8). These isolates were examined in more detail according to host family and species (Figures 8-11). The 14 moth isolates were recovered from 3 Lepidopteron families: Plutellidae, *Plutella xylostella* (n=7); Choreutidae, *Anthofilia fabriciana* (n=5); Noctuidae, *Helicoverpa armigera* (n=1) and *Agrotis ipsilon* (n=1). These isolates were found in both major phylogenetic clusters, but in order to examine host specificity in more detail a separate tree based on concatenated sequence was constructed containing only these isolates and colour coded according to host family using the termite isolate (T01) as an outgroup (Fig. 9). This tree reveals that the two major groups generally correspond to a division between the *Choreutidae* and the *Plutellidae*, although the two isolates corresponding to the Family *Noctuidae* are in separate groups. The exceptions are strains M07 and M13 which belong to *Choreutidae* and *Plutellidae* respectively, but cluster in the "wrong" groups. An inspection of the polymorphic sites reveals evidence of mosaicism, particularly within the genes *groEL* and *tpiA* (Fig. 10). For *tpiA* strains M06 and M22 have acquired short stretches of DNA at the extreme 5' end of the sequence corresponding

to sequence from the other cluster. In *groEL*, the mosaic structure is more complex, but strains M13 and M05 generally correspond to one cluster at the 5'end and the other at the 3' end of the sequenced allele.

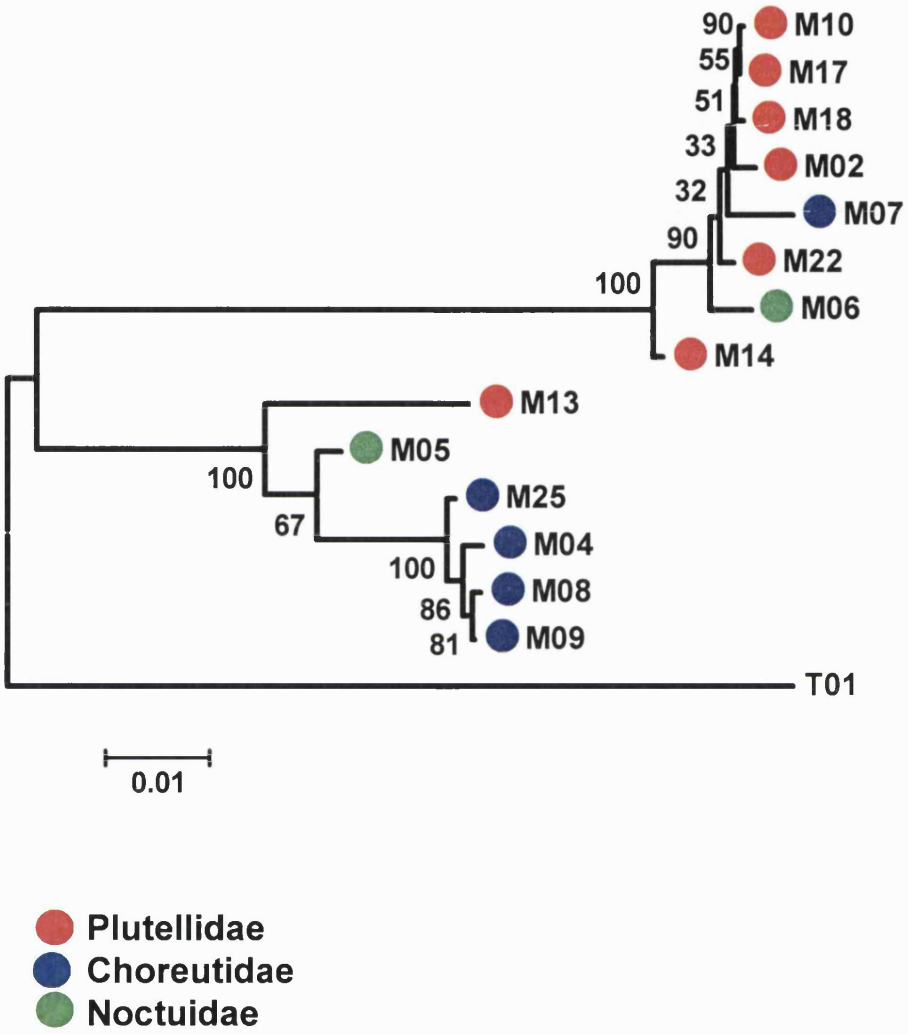



**Figure 9** Neighbour-Joining tree based upon concatenated alignment of five MLSA genes for *Wolbachia* spp. isolated from 14 Lepidoptera. Each strain is coded according to its invertebrate host; M = moth; T = termite. The tree has been colour coded according to Lepidopteron Family; Plutellidae (red), Choreutidae (blue) and Noctuidae (green). Strain T01 has been selected as an outgroup.

***groEL***

```
                    1111111111 1111111111 2222222222 3333333333 3333444444 44444444
     111222333 4445556778 0000111222 3345577789 1234455789 0123344566 7899000122 33567888
    6236167349 1890173284 0356016249 6740314798 9874728324 0223625406 8403358469 25357039
#M02 GTAAAGCATT GGGTCGCAAA TCCTCGTCAT GCGTCCAAGT GATATCTAGA ACTGTTAATA ATTCGTTGAC CCGAATAT
#M06 .......... .......... .......... .......... .......... .......... .......... ........
#M07 ...G...... .......... .......... .......... .........G .......... .......... ........
#M10 .......... .......... .......... .......... .......... .......... .......... ........
#M14 .......... .......... .......... .......... .......... .......... .......... ........
#M17 .......... .......... .......... .......... .......... .......... .......... ........
#M18 .......... .......... .......... .......... .......... .......... .......... ........
#M22 .......... .......... .......... .......... .......... .......... .......... ........
#M13 ...G...... C........C C.T...CTTC TGAC.AGGAC ATCGGTCG.. GTC..AGGCG GACTT..AGT AAAGGGTC
#M05 ...G...... .......... .A.GT.CT.. ..A.TA.... ....G....G GTCACG..A. GACTTCGAGT AAAGGGTC
#M04 AATGGAT.CA CAACTATGGC .....ACTTA TGACTA.GAC ATCGGTC.A. GTCACG..A. GACTTCGAGT AAAGGGTC
#M08 AA.GGATC.A CAA.TATGGC .A.GTACTTA TGACTA.GAC ATCGGTC.A. GTCACG..A. GACTTCGAGT AAAGGGTC
#M09 AA.GGATC.A CAA.TATGGC ...GTACTTA TGACTA.GAC ATCGGTC.A. GTCACG..A. GACTTCGAGT AAAGGGTC
#M25 AA.GGAT..A CAA.TATGGC ...GT.CTTA TGACTA.G.C ATCGG.C... GTCACG..A. GACTTCGAGT AAAGGGTC
```

***tpiA***

```
                11111111 1111122222 2222222333 333333333
     111233344 5801133446 7889911256 6667889122 334467889
    9018105689 4621959348 3032703981 2688254279 122532047
#M04 AGGTCCAAGC GGTTCTATCT AAAGCTGAGG GTGCCATTGA GAAATCCCG
#M08 .......... .......... .......... .......... .........
#M09 .......... .......... .......... .......... .....T...
#M05 ...C...... .......... .......... .......... .....T...
#M25 ...C...... .......... .......... .......... .........
#M13 ........AT A....G.C.C .....G...C ACTT....AG .GTG.....
#M06 ........AT AACAG.G... GGGAG.ACA. AC..A.CCA. A...CT.AA
#M22 .....TGGAT AACAG.G... GGGAG.ACA. AC..AGCCA. A...CT.AA
#M02 GAACTTGGAT AACAG.G... GGGAG.ACA. AC..AGCCA. A...CT.AA
#M10 GAACTTGGAT AACAG.G... GGGAG.ACA. AC..AGCCA. A...CT.AA
#M14 GAACTTGGAT AACAG.G... GGGAG.ACA. AC..AGCCA. A...CT.AA
#M17 GAACTTGGAT AACAG.G... GGGAG.ACA. AC..AGCCA. A...CT.AA
#M18 GAACTTGGAT AACAG.G... GGGAG.ACA. AC..AGCCA. A...CT.AA
#M07 GAACTTGGAT A..AG.G.T. GGGAG.ACA. AC..A.CCA. A...CTAAA
```

**Figure 10** Evidence of mosaicism within *groEL* and *tpiA* resulting from horizontal transfer between Lepidoptera.

This analysis provides evidence for some host adaptation on the level of host Family, as the isolates from the *Choreutidae* are generally clustered in one group and the isolates from the *Plutellidae* in another. However, there is also clear evidence of recombination between isolates recovered from different host Families and corresponding to different groups. There is also evidence for recombination from an analysis of the isolates recovered from parasitic wasps. Four parasitic wasp (Hymenoptera) Families are represented; Braconidae, *Macrocentrus grandii* (n=1), *Asobara tabidia* (n=2); Ichneumonidae, *Bracon hebetor* (n=1), *Phytodietus niger* (n=1), *Campoletis sonorensis* (n=1); Pteromalidae, *Roptrocerus xylophagorum* (n=1); unidentified sample (n=1). A concatenated tree for these isolates is given in Figure 11. Inspection of the polymorphic sites in strains PW01, PW04 and PW15 shows a mosaic structure when compared to the outgroup sequence T01 (Fig. 12), as the differences between strains PW04/PW15 and strain PW01 are clustered towards the 5' end of this gene.

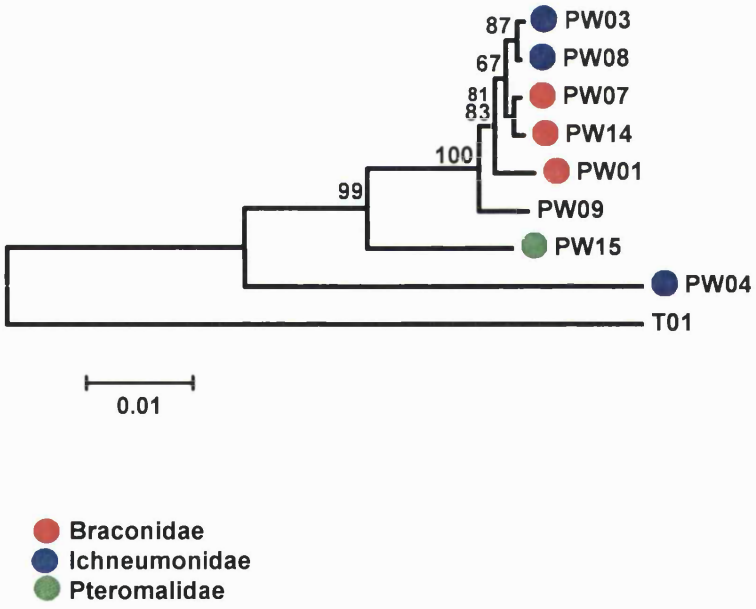



**Figure 11** Neighbour-Joining tree based upon concatenated alignment of five MLSA genes for *Wolbachia* spp. isolated from 8 Hymenoptera. Each strain is coded according to its invertebrate host; PW = parasitic wasp; T = termite. The tree has been colour coded according to Lepidopteron Family; Braconidae (red), Ichneumonidae (blue) and Pteromalidae (green). Strain PW09 has not been colour coded due insufficient evidence as to its classification. Strain T01 has been selected as an outgroup.

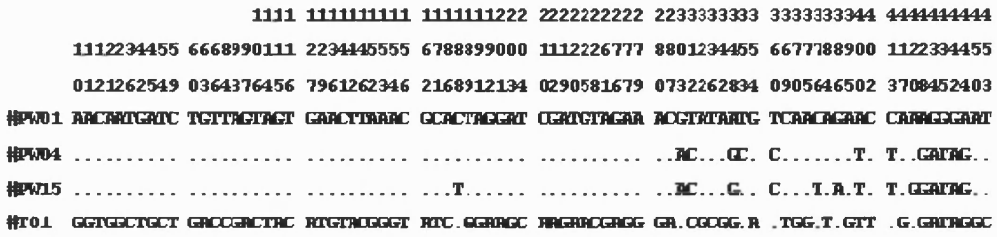



**Figure 12** Evidence of mosaicism within *trmD* resulting from horizontal transfer between the Order Isoptera (T01) and Hymenoptera (PW01, PW04, PW15).

### 3.3.5 *wsp* host incongruence

The *wsp* tree is by far the most divergent of the six loci sequenced (Fig. 7f), however due to only 26 instead of 44 host strains being available for analysis, it is difficult to compare *wsp* with the other gene loci. Nevertheless, it is clear that this gene also provides evidence of a decoupling of strain genotype with host family. Nine Lepidopteron moths were included in this analysis, including six Plutellidae. Four of these strains cluster together with the two other Lepidopteron butterfly species (BF01 & BF02; Family Hesperiidae & Pieridae, respectively), however the other two Plutellidae strains M13 and M14 are located distantly on the tree from the main cluster.

These data also provide evidence concerning diversification within the isolates recovered from laboratory stocks of *Drosophila* species. Strains F02, F04 and F05 were recovered from the same original stock culture of *Drosophila simulans* Riverside which was distributed to the University of Dundee, Imperial College London and the University Ioannia (Greece) respectively. These three strains cluster at all loci, with the exception of *trmD* where F02, F03 and F05 cluster together, and at *wsp* where F02, F03, F04 and F05 cluster together.

### 3.3.6 Phylogeny of *Wolbachia* genus supergroups

In order to compare the current data set with the previous assignments of *Wolbachia* isolates to supergroups, the data for *groEL* and *gltA* generated by Baldo *et al.* was retrieved from GenBank (Baldo *et al.*, 2005a). Alignments of these data with the data generated in this project allowed assignments of the isolates to known supergroups A and B (Table 9).

| Host species name | Accession Number | | Supergroups |
|---|---|---|---|
| | *gltA* | *groEL* | |
| *Camponotus sayi* | DQ266395 | DQ266398 | A |
| *Camponotus vafer* | DQ266396 | DQ266397 | A |
| *Protocalliphora sialia* | DQ266412 | DQ266419 | A |
| *Acraea eponina* | DQ266520 | DQ266417 | B |
| *Drosophila innubila* | DQ266411 | DQ266418 | B |
| *Nasonia giraulti* | DQ266527 | DQ266414 | B |
| *Teleogryllus taiwanemma* | DQ266529 | AB002286 | B |

**Table 9** Host species and accession numbers selected from Baldo *et al.*, (2005a).

Trees were constructed which included the data of Baldo *et al.*, (2005a) for *gltA* and *groEL* and these confirmed that the two main groups correspond to supergroups A and B (Fig. 13 & 14). As discussed above, strains B03 and PW04 correspond to one group at *gltA* (confirmed as supergroup A by comparison with the data of Baldo) and the other group (supergroup B) at *groEL*. This therefore indicates that recombination between *Wolbachia* isolates is not just limited to within each of the *Wolbachia* supergroups, but can occur between *Wolbachia* supergroups A and B. Isolate SM26, the only arachnid sample included in the study, clustered with supergroup B, complementing previous studies on spider mites and a number of Australian spiders (Breeuwer and Jacobs, 1996; Rowley *et al.*, 2004; Weeks and Breeuwer, 2001). Isolate T01, the only termite sample included in this study, did not cluster with either of the two supergroups; it is instead placed between the two groups on both trees. A Neighbour-Joining tree was constructed for *groEL* with sequences from further *Wolbachia* studies to determine whether T01 should be

classed with *Wolbachia* supergroup F or H (Bordenstein and Rosengaus, 2005; Casiraghi *et al.*, 2005). Figure 15 shows strain T01 clustering with sequence representing *Wolbachia* supergroup F. However, this could not be confirmed by comparison with the *gltA* sequences which appeared to be diverse and gave ambiguous alignments (data not shown).

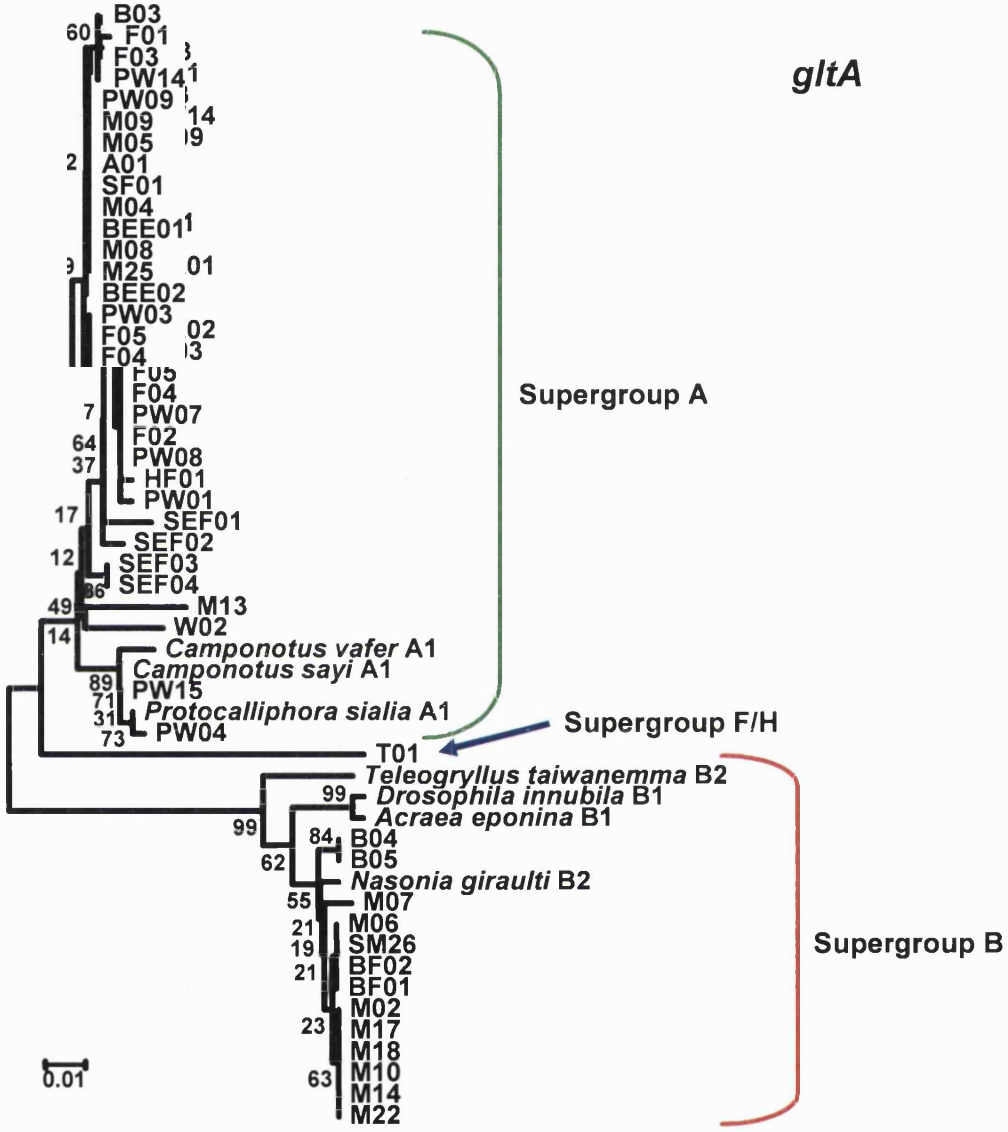



**Figure 13** Neighbour-Joining trees for *gltA*, identifying *Wolbachia* supergroups A, B and F/H (n=44). Each strain is coded according to its invertebrate host; A = ant; B = beetle; BEE = bee; BF = butterfly; F = fly; HF = hoverfly; M = moth; PW = parasitic wasp; SEF = stalk-eyed fly; SF = saw-fly; SM = spider mite; T = termite; W = wasp.

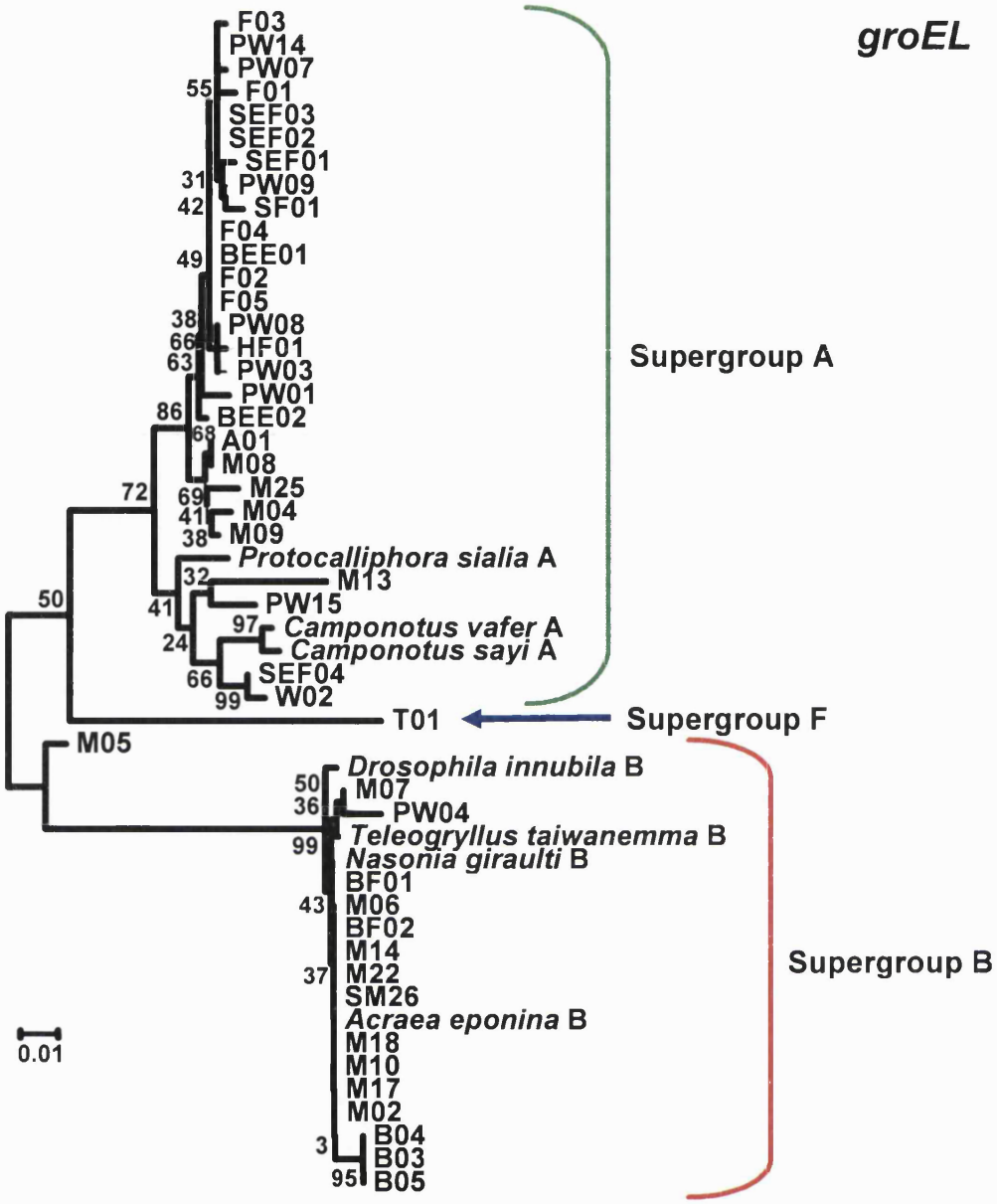



**Figure 14** Neighbour-Joining trees for *groEL*, identifying *Wolbachia* supergroups A, B and F (n=44). Each strain is coded according to its invertebrate host; A = ant; B = beetle; BEE = bee; BF = butterfly; F = fly; HF = hoverfly; M = moth; PW = parasitic wasp; SEF = stalk-eyed fly; SF = saw-fly; SM = spider mite; T = termite; W = wasp.

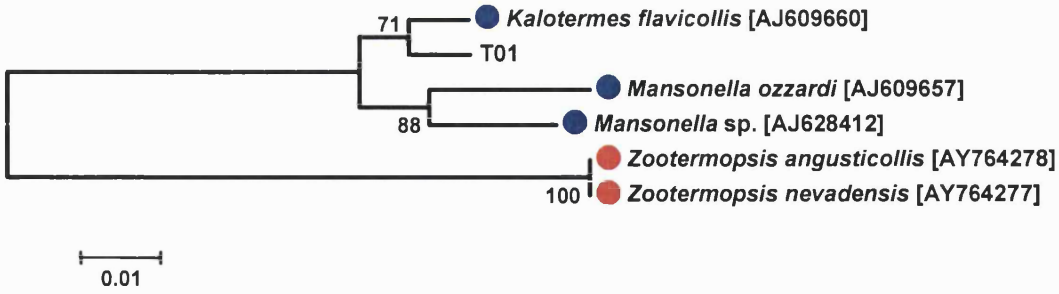



**Figure 15** Neighbour-Joining tree for *groEL* identifying isolate T01 as a member of *Wolbachia* supergroup F. Isolates from *Wolbachia* supergroups F and H are coloured blue and red respectively.

3.3.7 Biogeographical adaptation within *Wolbachia* spp.

The analysis so far has compared the trees for different loci, and provided evidence of recombination between isolates recovered from different host families and between supergroups. Despite frequent recombination, a close examination of the isolates from different moth Families provides evidence for some host specificity, as isolates recovered from the *Choreutidae* correspond to supergroup A whilst isolates from the *Plutellidae* correspond to supergroup B. However, a further question concerns to what extent the diversity of *Wolbachia* reflects geographical source independent of host species.

To address the extent of geographical structuring of *Wolbachia* spp., the five loci concatenated Bayesian tree was colour coded (Fig. 16) according to geographical source. In total, six categories were coded for; Europe (divided into UK and continental Europe); Asia; Americas; laboratory cultures; origin unknown. There is little obvious pattern, and isolates from a single source are clustered within both supergroups. Although the power of the dataset is somewhat limited for this analysis, this indicates that there is little or no geographic structure to the origin of *Wolbachia* spp. on a global scale. Strains originating from the UK, continental Europe, Asia and the Americas are each observed within the clusters which correspond to *Wolbachia* supergroups A and B. Only one possible example of

geographical structure was observed with the dataset, within the five Asian invertebrate samples. Isolates M17, M18 and M22 were collected from Pahang, Malaysia in 2001 and are observed within *Wolbachia* supergroup B. However, isolates SEF04 and T01 were collected from Thailand in 2004 and are observed in *Wolbachia* supergroup A. Furthermore, all the Malaysian *Wolbachia* were recovered from *Plutella xylostella* (the Diamondback moth) which is likely to account for the clustering of these isolates rather than the fact they were all recovered from the same location.

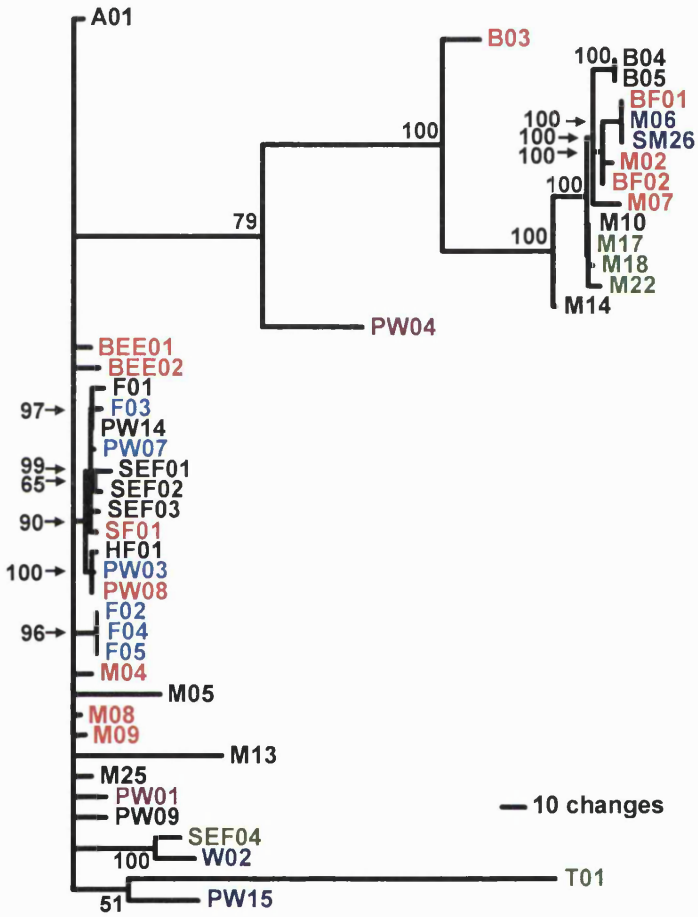



**Figure 16** Neighbour-Joining tree based upon concatenated alignment of MLSA loci for *Wolbachia* spp. (n=44). Each strain is firstly coded according to its invertebrate host; A = ant; B = beetle; BEE = bee; BF = butterfly; F = fly; HF = hoverfly; M = moth; PW = parasitic wasp; SEF = stalk-eyed fly; SF = saw-fly; SM = spider mite; T = termite; W = wasp; and secondly to the geographical location of the insect host species; red = UK; blue = continental Europe; green = Asia; purple = Americas; light blue = laboratory cultures; black = unknown location.

### 3.3.8 Evidence for recombination

To examine the extent of recombination within the population, the coalescence approach of Fearnhead & Donnelly (2001) was performed for each of the MLSA loci and *wsp* (Fearnhead and Donnelly, 2001). This approach is used to calculate the population-scaled recombination parameter ρ (Table 10), and has been widely used to examine MLST datasets (Perez-Losada *et al.*, 2006). The estimates of ρ range from 0 (the minimum value, observed for *atpA*, *gltA* & *trmD*) to 7.071 (*wsp*, n=26). Only two MLSA loci (excluding *wsp*) show evidence of recombination, *groEL* and *tpiA* (ρ = 3.03), which is consistent with previous analysis in which these genes are observed under greater selective pressures. Furthermore, a higher level of recombination was observed with *wsp*, consistent with the gene encoding an outer surface protein which is expected to be under greater diversifying selection. The mean ρ value for the MLSA loci (ρ = 1.21) is comparable to that observed in the human pathogenic species *Burkholderia pseudomallei, Staphylococcus epidermidis, Campylobacter jejuni, Enterococcus faecium* and *Escherichia coli*, (Perez-Losada *et al.*, 2006). Although the estimate of ρ is low, there is good evidence for extensive recombination in each of these species, suggesting that LDhat is a very conservative test. It is indeed surprising that *Wolbachia* appears to be in the same range as these human pathogens regarding recombination rates, as it might be imagined that the rate of migration and mixed infection might be much higher for human pathogens.

| Loci | LDhat estimate of ρ |
|---|---|
| *atpA* | 0 |
| *gltA* | 0 |
| *groEL* | 3.030 |
| *tpiA* | 3.030 |
| *trmD* | 0 |
| *wsp** | 7.071 |

**Table 10** Estimations of the recombinational parameter ρ within each MLSA loci and *wsp* using LDhat (n=44, *n=26).

The Sawyer's Runs Tests were also used to determine the extent of recombination within the *Wolbachia* dataset (Sawyer, 1989). This test has been widely used to infer recombination events within aligned sequences by determining if regions of sequence pairs have more consecutive identical polymorphic sites in common than would be expected by chance (see Chapter 1). The results of this test are shown in Table 11.

| Loci | Method of analysis | |
|---|---|---|
| | SSCF (P-value) | SSUF (P-value) |
| *atpA* | 19246 (P > 0.05) | 4832238 (P > 0.05) |
| *gltA* | 50770 (P > 0.05) | 11462362 (P > 0.05) |
| *groEL* | 315673 (P > 0.05) | 25615004 (P > 0.05) |
| *tpiA* | 51487 (P < 0.05) | 11714185 (P < 0.05) |
| *trmD* | 43409 (P > 0.05) | 11714185 (P > 0.05) |

**Table 11** Summary of the recombinational exchange using the Sawyer's Runs Test.

Neither the SSCF nor SSUF statistics revealed significant "mosaic" structure (a non-random distribution of polymorphisms) within any of the MLSA loci with the exception of *tpiA* ($P < 0.05$). Sawyer's Runs Test examines the rate at which recombination has effected the distribution of polymorphisms within aligned sequences, it is therefore insensitive to recombination events which have resulted in the replacement of whole alleles. Consequently, *groEL*, and perhaps to a much lesser extent *atpA, gltA* and *trmD*, have undergone large-scale replacements of the entire alleles. Such wholesale replacement of alleles is evident from the phylogenetic inconsistencies noted in PW14 and B03 (as discussed above). As the current evidence suggests that most replacements are large enough to encapsulate the whole gene, this suggests a larger role for transduction (phages) as a vehicle for recombination. As mentioned earlier in this chapter, *Wolbachia* spp. contain large amounts of mobile DNA, which is believed to have been introduced into the genome via phage, given that three prophage elements are present (Wu *et al.*, 2004). With

the large number of potential invertebrate hosts and the occurrence of multiple infections (Jamnongluk *et al.*, 2002), the potential for phage to move between *Wolbachia* spp. is great. However, it is also clear that from the examples of gene mosaicism discussed earlier in this chapter, that occasional recombination events only replace portions of the alleles. The non-significant P value from Sawyer's Runs Test indicates that this tends to be the exception rather than the rule.

The Homoplasy Test (Maynard Smith and Smith, 1998), which analyses the true homoplasies among informative, synonymous polymorphic sites, was used to determine the extent of recombination amongst the *Wolbachia* MLSA data (Table 12). Homoplasy ratios below zero were observed for all MLSA loci, in addition to non-significant probabilities ($P > 0.05$), indicating significantly fewer homoplasies than would be expected under free recombination. Therefore, this test also indicates that recombination has not been completely pervasive in the sequences.

| Loci | Homoplasy Test | |
|---|---|---|
| | P-value | H Ratio* |
| *atpA* | P = 1 | -0.157 |
| *gltA* | P = 1 | -0.078 |
| *groEL* | P = 1 | -0.099 |
| *tpiA* | P = 0.541 | -0.037 |
| *trmD* | P = 1 | -0.061 |

**Table 12** Homoplasy Test results for MLSA loci (n=44). *Homoplasy Ratio.

To determine the degree of linkage disequilibrium between alleles among the MLSA loci, the Index of Association ($I_A$) was performed (Smith *et al.*, 1993). The $I_A$ for all 44 isolates was found to be 0.932, which is significantly higher than zero ($P<0.001$), indicating that *Wolbachia* spp. has a clonal population structure where recombination is limited. Therefore, despite the evidence for recombination from small number of phylogenetic inconsistencies, and from visual inspection of polymorphic sites

revealing mosaic structure, the rates of recombination are not so high as to be detected by LDhat (which gave low values of ρ), Sawyer's Runs test, the Homoplasy test or the Index of Association.

### 3.3.9 Splits decomposition

As a final test for recombination, Splits decomposition analysis was performed (Bandelt and Dress, 1992). The analysis recovered similar phylogenetic clusters as the other more traditional phylogenetic approaches (Fig. 17); the splits decomposition networks divided the strains into supergroups A and B. The splittability index, the goodness of fit to the data, is presented in Table 13; these estimates range from 57.11 % (*groEL*) to 90.35 % (*trmD*). Where no reticulations are present, the graph presents as a clonal structure. Conversely, reticulations in splits graphs are indicators of conflicts in the data caused by recombination. All loci, with the exception of *atpA*, show reticulate structure primarily localised near the tips, indicating more recombination within groups than between groups.

Within *tpiA*, the presence of reticulations indicates that recombination has occurred only within *Wolbachia* supergroup B and not in supergroup A, or between the supergroups. Reticulations within each supergroup are observed within *gltA*, but again not between the groups which are very distinct and separated by a long branch. Conversely, *groEL* and *trmD* possess significant reticulations within the base of the graph rather than at the edges, indicating recombination has occurred between the two *Wolbachia* supergroups. All loci, with the exception of *atpA*, present isolate T01 on a long branch, intermediate to the two supergroups although typically closer to supergroup A. This is consistent with the trees discussed earlier, in that it is a member of a more diverse *Wolbachia* supergroup, probably supergroup F.

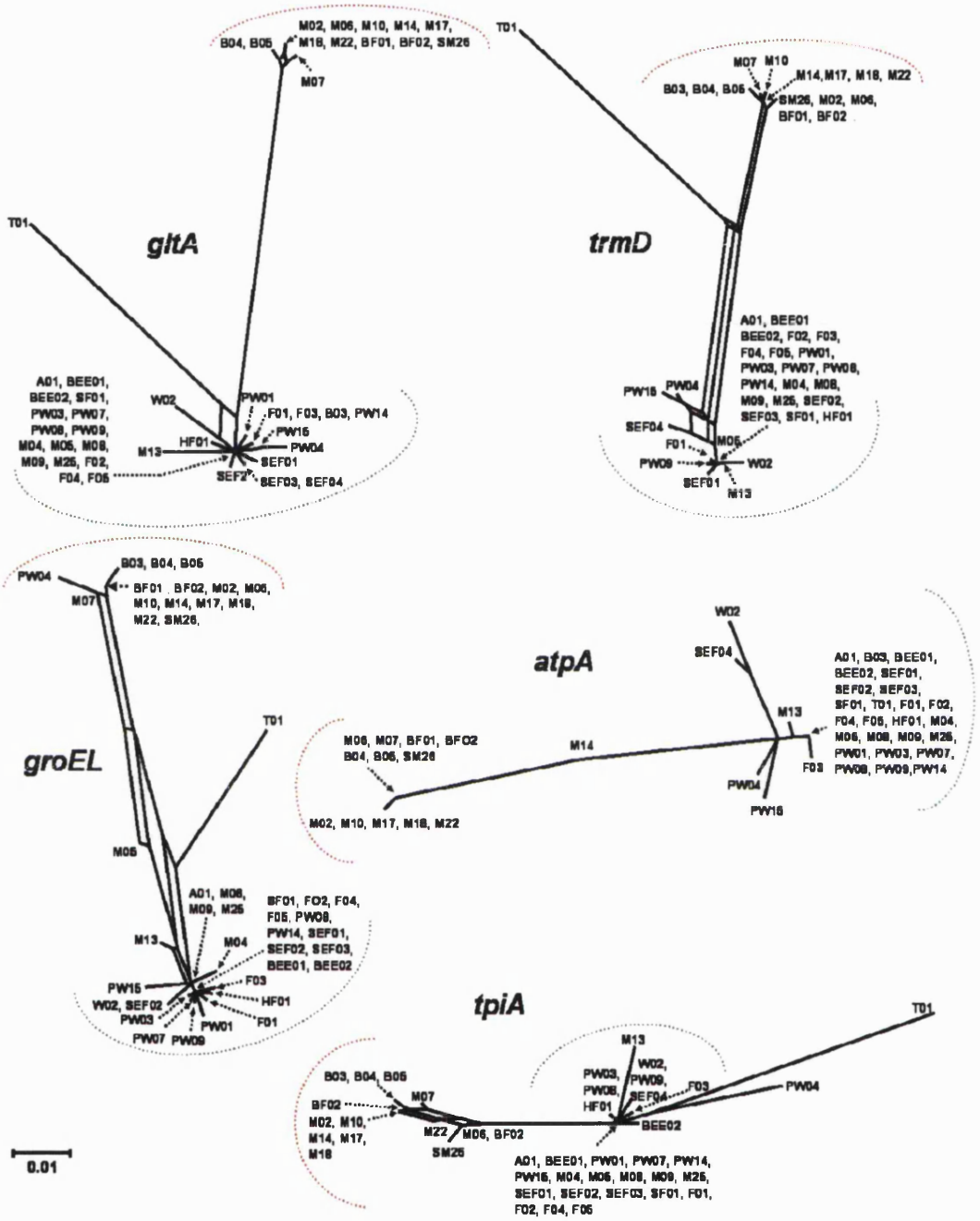



**Figure 17** Splits decomposition analysis of each MLSA loci (n=44). *Wolbachia* supergroups A and B are highlighted by green and red brackets, respectively.

| Loci | Splittability Index |
|---|---|
| *atpA* | 88.55 % |
| *gltA* | 77.44 % |
| *groEL* | 57.11 % |
| *tpiA* | 65.42 % |
| *trmD* | 90.35 % |

**Table 13** Splittability Index values for each MLSA locus (n=44).

3.3.10 Clonal diversification

Allelic profiles were assigned on the basis of the nucleotide sequence of each locus, and the structure of the population examined using eBURST (Fig. 18). Forty unique genotypes (STs) were determined, the majority of which were represented by only a single isolate (90 %), while four are represented by at least two isolates. No dominant clone was observed, instead four STs occurred with equal frequency (ST23, ST24, ST28 and ST31), all of which possessed two isolates. The 40 STs were divided by eBURST into one clonal complex, two triplets, two doublets and fourteen singletons. From a "snapshot" of the population (group definition 0/5 alleles in common) one clonal complex is evident and corresponds to *Wolbachia* supergroup A in addition to the majority of the singletons. *Wolbachia* supergroup B however does not follow a model of radial diversification. Composed of 15 STs and isolates, the founder clone of the complex is ST9 (CC9), however with only two SLVs and five DLVs linked to this isolate, it should be regarded as a minor clonal complex. Of the STs which include more than one isolate, both ST28 and ST31, represent *Wolbachia* spp. isolated from identical hosts, *Drosophila simulans* and *Plutella xylostella* respectively. Interestingly, two STs are found in more than one host species, with ST23 observed in two Coleoptera, B04 & B05, both of which were isolated from different species, *T. confusum* & *T. castaneum* respectively, and ST24 found in two Lepidoptera, BF01 & M06, isolated from an unidentified Hesperiidae and *Agrotis ipsilon* respectively.

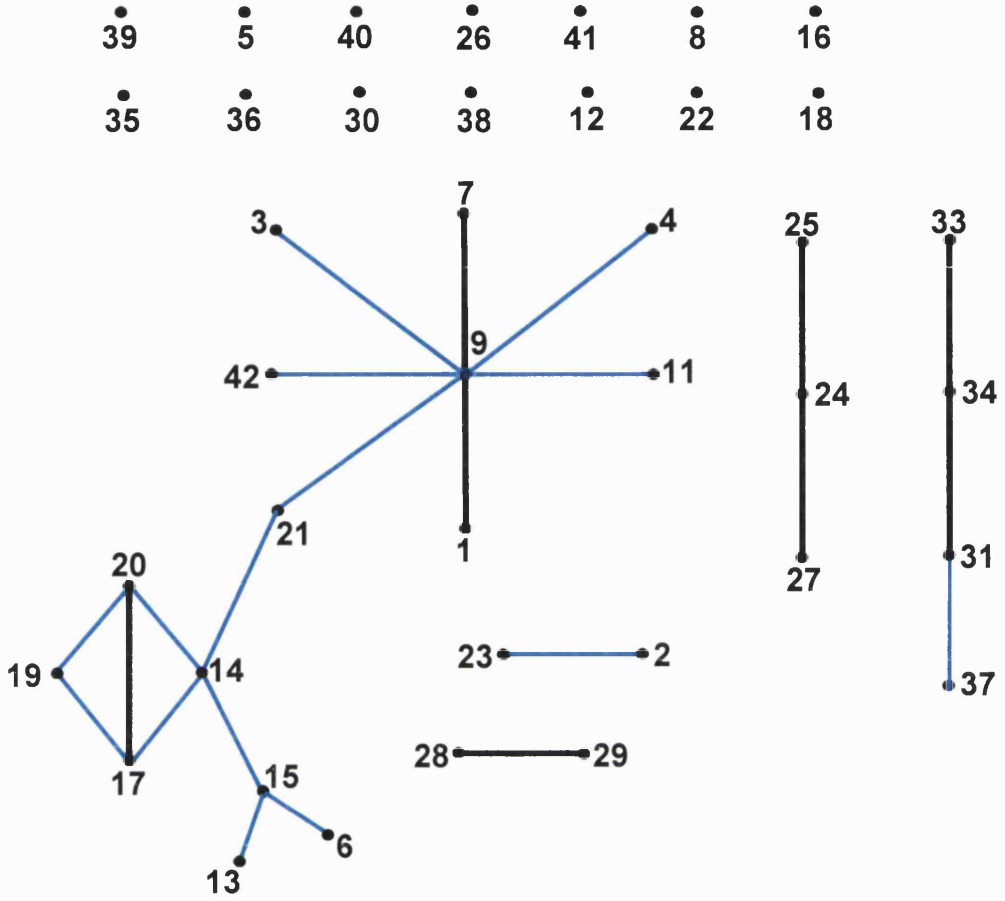



**Figure 18** A population "snapshot" showing the clusters of linked STs and unlinked STs for 44 *Wolbachia* spp. using nucleotide sequence data (0/5 alleles in common). The predicted clonal ancestor is shown in blue, SLVs are shown as black lines and DLVs as light blue lines. The sizes of the circles that represent each ST indicate their prevalence within the population.

### 3.3.11 eBURST based on translated alleles

As eBURST was designed to examine the patterns of descent within single clonal complexes of human pathogens, the *Wolbachia* data is too diverse for the analysis to provide much information. In order to address this, allelic profiles were re-assigned on the basis of their amino-acid, rather than nucleotide sequence, and examined by eBURST (Fig. 19). By removing all the synonymous changes, this method essentially "slows down" the diversification within the sample and it thus becomes possible to examine the patterns over evolutionary descent. A total of thirty-six STs were noted, with 86.1 % (n=31) of all STs represented by only a single isolate, and 13.9 % of all STs (n=5) represented by the minimum of two isolates. In order to cross reference the ST assignments based on nt- and aa- alleles, a table is presented in Appendix A. Two clonal complexes are noted using the translated alleles, each with 2 subgroups, along with seven singletons. Using the *Wolbachia* supergroup allocations observed with *groEL* (Fig. 14), the two clonal complexes were found to be consistent with the two supergroups. The largest clonal complex, aa-CC4, consists exclusively of *Wolbachia* supergroup A isolates, and the founder of this complex is aa-ST4 (bootstrap support 98 %). This ST corresponds to four nt-STs (STs 3, 26, 28, 29) indicating that a large proportion of the diversity is synonymous. In addition, aa-ST4 (n=5) is observed in 3 different invertebrate species (Table 14) across 2 insect Orders, Hymenoptera (BEE01, PW07) and Diptera (F02, F04, F05). The finding that this ST is a generalist regarding host choice is consistent with its assignment as the founder of supergroup A. Interestingly, isolates F02, F04 and F05 are all derived from the same *Wolbachia pipientis* laboratory strain used by the University of Dundee, Imperial College London and University of Ioannia, respectively. A small number of synonymous changes are detected in these strains, suggesting that they have diversified since the establishment of the stock cultures. For example, they differ at *trmD* at nucleotide level by a synonymous mutation (T > G base substitution). Also from aa-CC4, aa-ST9 contains two isolates, both of which belong to the Order Coleoptera, but are from different hosts species (B04, *T. confusum*; B05, *T. castaneum*), and aa-ST1 contains two isolates, A01 which belongs to the Order Hymenoptera and M08, *A. fabriciana* a Lepidopteron moth. The second clonal complex, aa-CC23, is composed only of *Wolbachia* supergroup B isolates. The founder of aa-CC23 is ST23 and is composed of only one isolate, nt-ST25. Subgroup founders aa-ST27 and aa-ST28 and their SLVs were all isolated

from the same Lepidopteron species, *P. xylostella.* aa-ST21 is composed of two isolates, both of which belong to the Order Lepidoptera, but belong to different species; BF01, unidentified butterfly, M06, *A. ipsilon.* Only two singletons are present at protein level, aa-ST20 and aa-ST25, with the latter grouping with *Wolbachia* supergroup B. aa-ST20, isolate T01, is hypothesised to fall into either *Wolbachia* supergroup F or H and has been placed intermediate to the two groups on Figure 18.

| Isolate | Invertebrate host |
|---|---|
| BEE01 | unknown Hesperiidae |
| F02 | *Drosophila simulans* |
| F04 | *Drosophila simulans* |
| F05 | *Drosophila simulans* |
| PW07 | *Asobara tabidia* |

**Table 14** Five isolates observed in aa-ST4.

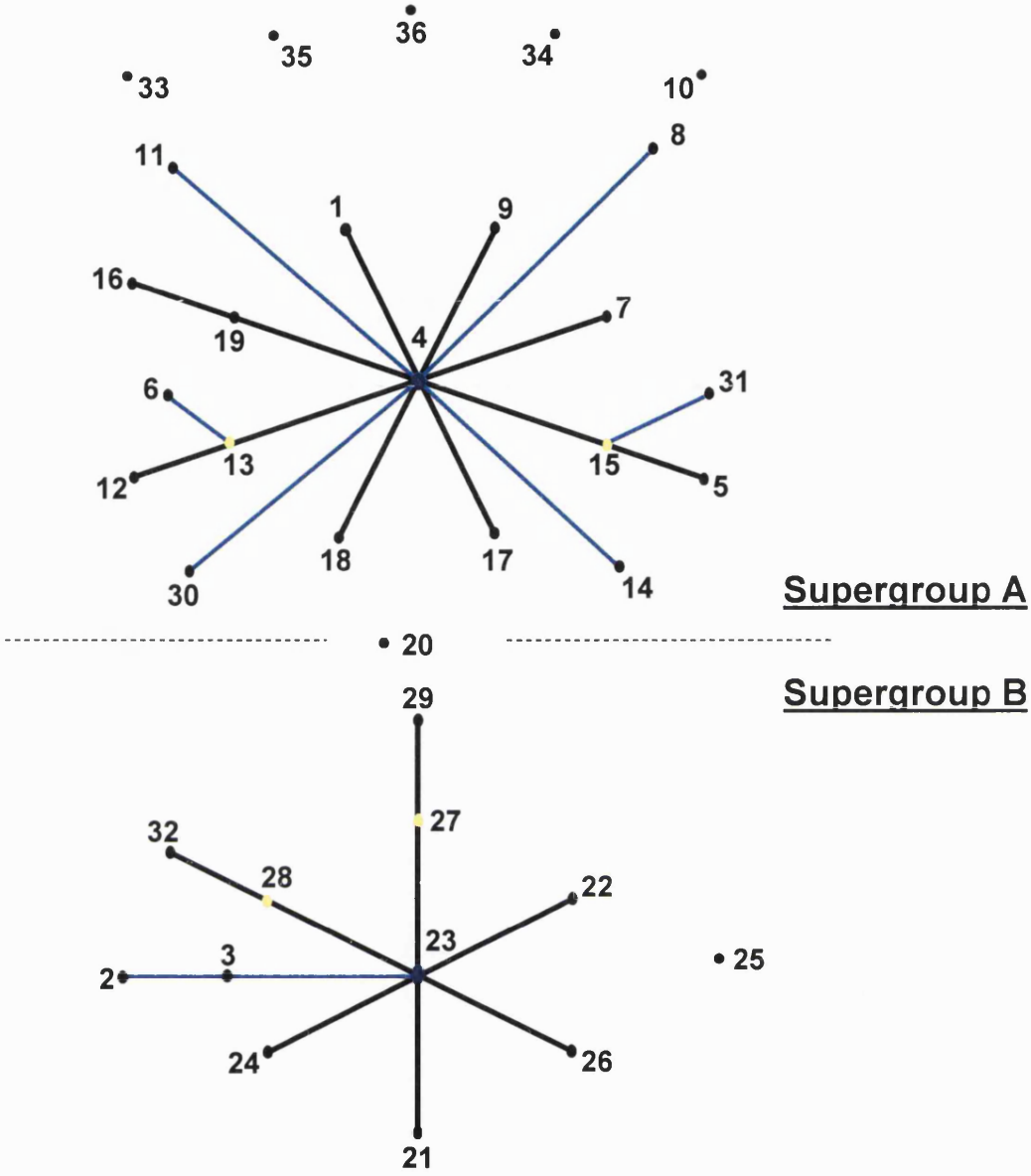



**Figure 19** eBURST applied to translated allelic profiles. A population "snapshot" showing the *Wolbachia* supergroup clusters of linked STs and unlinked STs for *Wolbachia* spp. using amino acid sequence data (0/5 alleles in common). The predicted clonal ancestors are shown in blue, the subgroup founders shown in yellow. SLVs and DLVs are shown as black and blue lines respectively. The sizes of the circles that represent each ST indicate their prevalence within the population.

## 3.4 DISCUSSION

*Wolbachia* are thought to be the most prevalent bacterial symbiont in insects, are vertically transmitted obligate symbionts and cause reproductive alterations in their hosts through a variety of mechanisms; cytoplasmic incompatibility, parthenogenesis, male killing, feminization (Bouchon *et al.*, 1998; Charlat *et al.*, 2003; Jeyaprakash and Hoy, 2000; Werren *et al.*, 1995a; Werren *et al.*, 1995b; Yen and Barr, 1971).

Although much is known about the reproductive manipulations of *Wolbachia*, little is understood of its population structure, or specifically the rate of recombination and the extent of host specialisation. Five MLSA genes were used in this study to examine the diversity and population structure within the *Wolbachia* genus, and to estimate the mechanisms involved in this diversity. This was supplemented with *wsp* data for a sub-sample of the strains. It has been speculated by comparison with other obligate endosymbionts of insects, that the rate of recombination should be very low in *Wolbachia* spp. This is because of the limited opportunities that different lineages would have to meet in the wild, owing to the inability to survive outside of the host environment and strict vertical transmission between hosts. As well as being consistent with low rates of recombination, such a lifestyle might also predict co-speciation between the host and bacterium as has been observed for *Buchnera* and Aphids (Wernegreen and Moran, 1999). This analysis therefore focused on evidence for recombination and host specialisation, as this provides clues as to the frequency of mixed infection and horizontal transfer between hosts in the wild.

Nucleotide sequence data for forty-four isolates was generated for five loci (*atpA, gltA, groEL, tpiA, trmD*). In addition, sequence data for twenty-six of these isolates was generated for *wsp*, the traditional gene of choice for the molecular identification of *Wolbachia*. Isolates spanning six insect Orders, from four continents were included in the analysis. These data provide the means to address the aims listed in Chapter 1 and these will now be discussed in turn:

Phylogenetic incongruence to host phylogenies

From the five MLSA loci sequenced, clustering according to host Family was observed, however the data generally did not distinguish between isolates from the same host family but different host species. The more limited data for *wsp* also revealed inconsistencies between clustering and host species. The sample contained the highest number of strains from Lepidoptera and Hymenoptera isolates, in particular, moths and parasitic wasps which allowed individual detailed studies on host congruence within individual species & Families. Although the topology of these isolates remains relatively stable across each of the MLSA loci, horizontal transfer has resulted in incongruence between individual hosts with mosaicism clearly evident. It is likely that the ancestral genome of these species was a mosaic from many different lineages. Although with a limited sample number (n=26), host incongruence was also observed within *wsp*. Because this gene encodes a proteins associated with the cell surface, it is likely to be involved in interactions with the host environment, and thus potentially plays a role in host specialisation. However, the inconsistencies between clustering and host species observed at *wsp* suggest that this gene probably does not play a dominant, or sole, role in host adaptation.

The MLSA data is consistent with the *Wolbachia* supergroups

*Wolbachia* supergroups were originally classified into taxonomic groups on the basis of clustering patterns in *ftsZ*-based phylogenetic trees (Bandi *et al.*, 1998; Lo *et al.*, 2002; Werren *et al.*, 1995b). Using data from recent studies which have focused on housekeeping genes (Bordenstein and Rosengaus, 2005; Casiraghi *et al.*, 2005), we were able to assign the strains to supergroups on the basis of the MLSA data. Comparisons using *gltA* and *groEL* confirmed that the two distinct clusters correspond to *Wolbachia* supergroups A and B. The isolate from a termite T01, probably corresponds to supergroup F at *groEL*. However, unambiguous assignment of this isolate is not possible with the current data set.

Baldo *et al.*, suggested that host-specificity is lower in *Wolbachia* supergroup B isolates than in other supergroups, which could be explained by a higher rate of horizontal transmission of bacteria between hosts (Baldo *et al.*, 2005a; Werren *et al.*,

1995b). The current data supports this view, with recombination clearly evident between isolates corresponding to supergroup B. Horizontal transfer within and between supergroups A and B has been noted previously (Baldo *et al.*, 2005a; Malloch and Fenton, 2005) and is strikingly apparent in the two isolates, B03 and PW04 as these strains cluster with supergroup A for *gltA* and supergroup B with *groEL*.

Lack of geographical structure within the *Wolbachia* population

With *Wolbachia* believed to infect 16.9 % – 76 % of terrestrial invertebrate species (Bouchon *et al.*, 1998; Jeyaprakash and Hoy, 2000; Werren *et al.*, 1995a), it undoubtedly has a global distribution. However, it is possible that there is significant geographical structure in the population owing to restricted migration between hosts, or restricted geographical ranges of the hosts themselves. Invertebrates may have limited geographic ranges and often are localised to a particular region due to food requirements and geographical boundaries. The limited evidence from the current data is not consistent with strong geographical structuring, although such an effect may be difficult to decouple from host association. It is clear, however, that isolates from any given geographical source can correspond to either supergroup A or B, suggesting that there is no structuring on the continent / supergroup level. This does not exclude the possibility however that there may be geographical structuring on a much finer scale, between strains associated with closely related hosts (see Chapter 4).

Evidence for recombination within *Wolbachia* spp.

Within *Wolbachia*, large-scale recombination events are not expected, due to the bacterium being vertically transmitted thus reducing the probability of two strains meeting. Recombination may have different selective outcomes for the bacteria. It may in some cases be adaptive if it results in the acquisition of genes which facilitate the exploitation of a new niche (such as a different host species). Alternatively it may be selectively neutral or deleterious. Other endosymbionts are thought to be undergoing a process of genome degradation due to a small population size and weakened purifying selection. It is possible that the high frequency of IS elements

observed in the *Wolbachia* genome also reflects weakened selection as they have a deleterious effect on the structure of the genome. In the same way, it is not safe to assume that a high frequency of recombination is necessarily of adaptive value to the bacterium.

The analysis provides evidence as to the significance of recombination in *Wolbachia* on a gene by gene basis. *atpA* is the most evolutionary conserved protein-coding loci, displaying the lowest levels in all recombination analyses. The splits decomposition analysis provides the most visual evidence of this as no reticulations are noted at *atpA*. This gene has not previously been used as a phylogenetic marker for *Wolbachia*, but this analysis suggests that, as a single gene marker, it should prove extremely useful. At the other end of the scale, recombination is perhaps most evident at *groEL*, which is also the most diverse locus. This locus was believed to undergoing positive selection in *Buchnera* due to its role in the folding of other proteins in the cell cytoplasm.

The finding of phylogenetic inconsistencies and significant mosaic structure provides unequivocal evidence of recombination in *Wolbachia*, thus confirming the recent report of Baldo *et al.*, (2005a). A particularly striking example is the base-perfect mosaic structure evident in Figure 8, which represents a recombination event both between supergroups (A and B) and between different host orders (Hymenoptera and Lepidoptera). The fact that this mosaic is base perfect suggests that it may have happened relatively recently, as there is no evidence of subsequent mutation.

However, despite these clear examples, it is unclear whether the data justify the rate of recombination being described as "pervasive" as it was described by Baldo *et al.*, (2005a) as there are plenty of signatures of clonal evolution evident in the data. First, with a few exceptions, all the gene loci agree on the two major groups, and the assignment of strains to "supergroups" has therefore not been fully compromised by free recombination. Secondly, many of the tests detected low rates of recombination, or failed to find evidence for any at all, which suggests that the mosaic structure represents the exception rather than the rule and were only noted through extensive visual inspection of the data. Thus *Wolbachia* probably occupy a

"middle ground" between freely recombining pathogenic species such as *Neisseria* and the strictly clonal *Mycobacterium*.

The fact that recombination occurs at all raises a number of interesting questions. Either recombination occurs within a single host, in which case the rate of mixed infection must be high, and the level of host specialisation low, or alternatively there may exist an unsuspected environmental reservoir outside of the host where *Wolbachia* might survive temporarily. Recent reports have suggested that the phylogeny of *Wolbachia* does not tightly correspond to host species, beyond the broad groupings of arthropods and nematodes described above, which implies that *Wolbachia* must utilise a combination of vertical and horizontal transmission (O'Neill *et al.*, 1992a; Rousset *et al.*, 1992b; Werren *et al.*, 1995a; Werren *et al.*, 1995b; Zhou *et al.*, 1998). The possibility of horizontal transmission remains contentious, however, as *Wolbachia* has never been observed to be free living, and it is unclear how this might occur.

One possible site whereby *Wolbachia* may meet outside of the host environment is on the surface of plants. The role of plants as potential environmental reservoir was discussed recently by Sintupachee *et al.*, (2006). It is known that *Wolbachia* can infect the mid-gut of insects, and it is therefore likely that pollinators "shed" *Wolbachia* cells on plants, which may then potentially infect the next host to visit the plant. Even if the cells lyse on the plants it is conceivable that *Wolbachia* DNA might find its way into another host, possibly via phage. Phage transduction is likely to represent the most efficient method of genetic transfer in *Wolbachia* (three prophage are noted in the *w*Mel genome (Wu *et al.*, 2004). Phage transduction tends to result in the replacement of larger pieces of DNA than transformation. This could help to explain why Sawyer's Runs test did not detect evidence of recombination, as the majority of events involve the replacement of whole alleles.

# CHAPTER FOUR

## EVOLUTION & MOLECULAR CHARACTERISATION OF *WOLBACHIA* ENDOSYMBIONTS IN THE ORDER ACARI

# 4.1 INTRODUCTION

## 4.1.1 Biology of spider mites

Spider mites are small (usually less than 1mm length) arthropods that may be brown, red, green or cream-colored. Commonly misidentified as insects, spider mites are members of the Class Arachnida, Order Acari. Unlike insects which have six legs and three body regions (head, thorax, abdomen), adult spider mites have eight legs and a single body region containing the gnathosoma and the idiosoma. In addition, they also lack antennae, compound eyes and wings, and have the ability to produce fine silk webbing (Fig. 1).

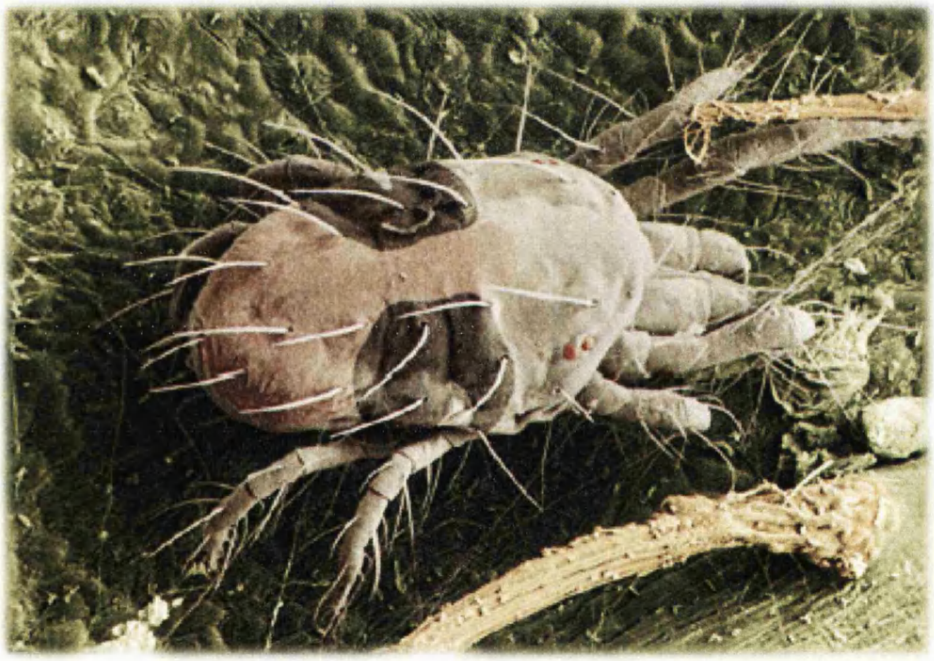


**Figure 1** Scanning electron micrograph of the two-spotted spider mite, *Tetranychus urticae,* feeding on a rose leaf (Kane and Ochoa, 2006).

### 4.1.2 Habitat and distribution of Acari

Approximately 45,000 species of mites have been described (Walter *et al.*, 1996), and are among the oldest of all terrestrial animals, with fossils known from the early Devonian, nearly 400 million years ago (Norton *et al.*, 1988). Mites are ubiquitous arthropods observed in a wide variety of habitats (terrestrial, marine and freshwater), environments (ranging from to nutrient rich to barren deserts) and climate regions (polar, temperate and tropical) (Bartsch, 2004; Hountondji *et al.*, 2005; Marshall and Convey, 2004; Walter *et al.*, 1996). Mites form complex relationships with the larger organisms on which they live. Firstly, birds and mammals are hosts to numerous species of parasitic mites (e.g. *Sarcoptes scabei*), as are many reptiles, insects and amphibians (Ford *et al.*, 2004; Heukelbach and Feldmeier, 2006; Proctor and Owens, 2000; Sammataro *et al.*, 2000). Secondly, many mite species also live and feed on plant flora and are therefore considered major economic pests. For example, the two spotted spider mite (*Tetranychus urticae*) attacks a wide range of garden plants, including many vegetables (e.g., beans, eggplant), fruits (e.g., apples, cucumber, strawberries) and flowers (Blindeman and Van Labeke, 2003; Kim and Lee, 2003; Maeda *et al.*, 2000; Park and Lee, 2005; White and Liburd, 2005). Finally, Phytoseiid mites (Family: Phytoseiidae) are predatory arachnids of spider mites (Croft and Jung, 2001). Often used as biological control agents of spider mite pests, these mites are much smaller than their counterparts at 0.5 – 0.8 mm long, and live in the soil or leaf litter surrounding plants (Croft and Jung, 2001).

As discussed in Chapter 3, *Wolbachia* supergroup G contains strains isolated from Australian arachnids and a fig wasp. Despite the fact that spider mites are more related to arachnids than they are to insects, studies with *ftsZ* have shown *Wolbachia* from spider mites are associated only with supergroup B (Breeuwer and Jacobs, 1996).

4.1.3 Spider mites utilise both sexual and asexual reproduction

In the animal kingdom, reproduction can occur by two mechanisms; either sexually or asexually (Fig. 2). Sexual reproduction is the favoured method of reproduction for many organisms. In sexual reproduction the formation and fusion of two different kinds of gametes (n) to form a zygote (2n) results in progeny with a somewhat different genetic constitution from either parental type and from each other. In this way, genetic variation (recombination) is introduced and maintained, creating genetic diversity within a population. Such variation is an advantage to the species if the environment is changing, because an offspring might be better adapted to the new environment than is either parent. In Lewis Carroll's *Alice in Wonderland*, the Red Queen must constantly run just to keep in the same place, a metaphor that has been used to explain the evolution of sex. According to the "Red Queen" hypothesis, species must constantly produce variability through sex and recombination just to keep apace of co-evolving species, including parasites, predators and competitors (Ochoa and Jaffe, 1999). This is achieved through sexual reproduction. Costs of this process include the need for two individuals to mate, courtship rituals, cost of males within the population and recombination. Asexual reproduction does not involve formation and fusion of gametes, instead results in progeny with an identical genetic constitution to the parent and to each other, i.e. clonal inheritance. This lack of variation is not a disadvantage as long as the environment remains stable. Organisms that reproduce asexually often have the advantage of being able to produce a large number of offspring within a limited amount of time. Only certain methods of asexual reproduction are found among animals, for example, regeneration and budding. Some eukaryotes have the ability to reproduce parthenogenetically, e.g., some species of plant, invertebrate (honey bees) and vertebrates (e.g. lizards and salamanders). Parthenogenesis is the development of individuals from unfertilised eggs, therefore the populations must be all female because there is no contribution from a male. Offspring may be capable of sexual reproduction, however, if that exists within the species. For example, aphids can be alternately parthenogenetic and then fully sexual. Aphids reproduce parthenogenetically in spring and summer when the food supply is plentiful. This enables them to reproduce very rapidly (and produce wingless offspring). Approaching winter with limited food supplies, they switch to sexual reproduction and deposit their fertilized large eggs on plants which begin a new generation of

asexually reproducing (winged) females the following spring (Mader, 1996). The exception of sexual reproduction resulting in haploid offspring is what is seen in social insects such as bees and wasps; haplodiploidy. In this situation, male offspring are haploid as they result from unfertilized eggs, and females are diploid from fertilized eggs. This process is arrhenotoky and results in sisters which are genetically 75% identical to each other as they all get the same haploid set from their fathers and either haploid set from their mothers. In thelytokous parthenogenesis, the second set of chromosomes comes not from sperm, but from one of the three polar bodies during meiosis.

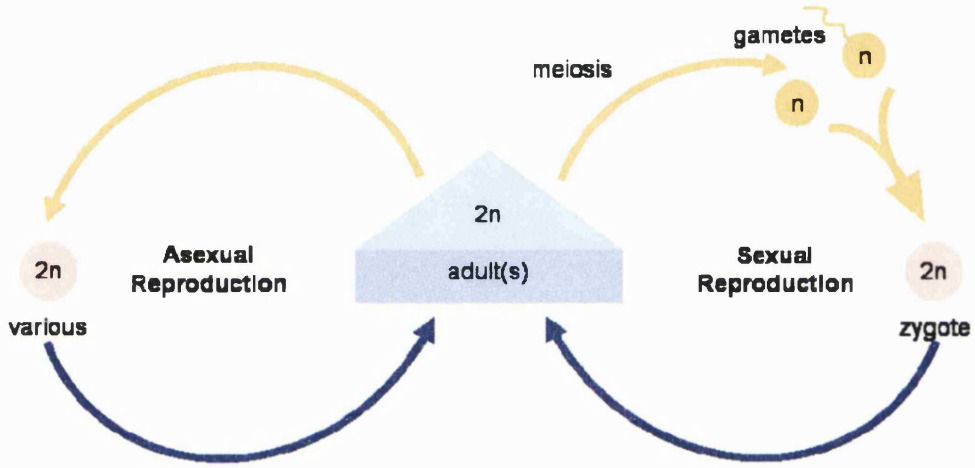



**Figure 2** Asexual versus sexual reproduction in animals. Modified from Mader *et al.*, (1996).

### 4.1.4 Spider mite life cycle

Spider mite development varies somewhat between species, however most aspects remain common across the Tetranychidae Family. The spider mite life cycle consists of four active stages; i) six legged larval stage, ii) & iii) eight legged nymphal stages, protonymph and deutonymph respectively, and iv) an eight legged adult stage (Fig. 3). Between each active stage is a quiescent developmental (chrysalis) stage that is sessile but physiologically active. During the egg and chrysalis stages, the mites remain attached to the plant surface, and do not move further than a few millimetres from their hatching place unless food conditions are

poor (Vala *et al.*, 2004). The time taken for the development of the egg to adult form varies and is dependent on the species of spider mite and the temperature conditions. Within *T. urticae* populations, eggs may hatch anywhere between two and fifteen days depending on temperatures, with populations spreading rapidly at optimum temperature of 30 – 32°C. Development is retarded below 12°C and above 40°C (Burnip *et al.*, 2006). Female spider mites survive for approximately 3 – 4 weeks and can lay between 70-100 eggs during one oviposit (Kane and Ochoa, 2006). Development from egg to adult takes up to 24 days, however during summer this can be as short as 2 – 3 weeks. As winter approaches, with shorter days, lower temperatures, and declining leaf condition, reproduction ceases and all spider mite stages except the females die. The adult females then move to their overwintering sites (*T. urticae* overwinters in soil) and lie dormant awaiting spring (Burnip *et al.*, 2006).

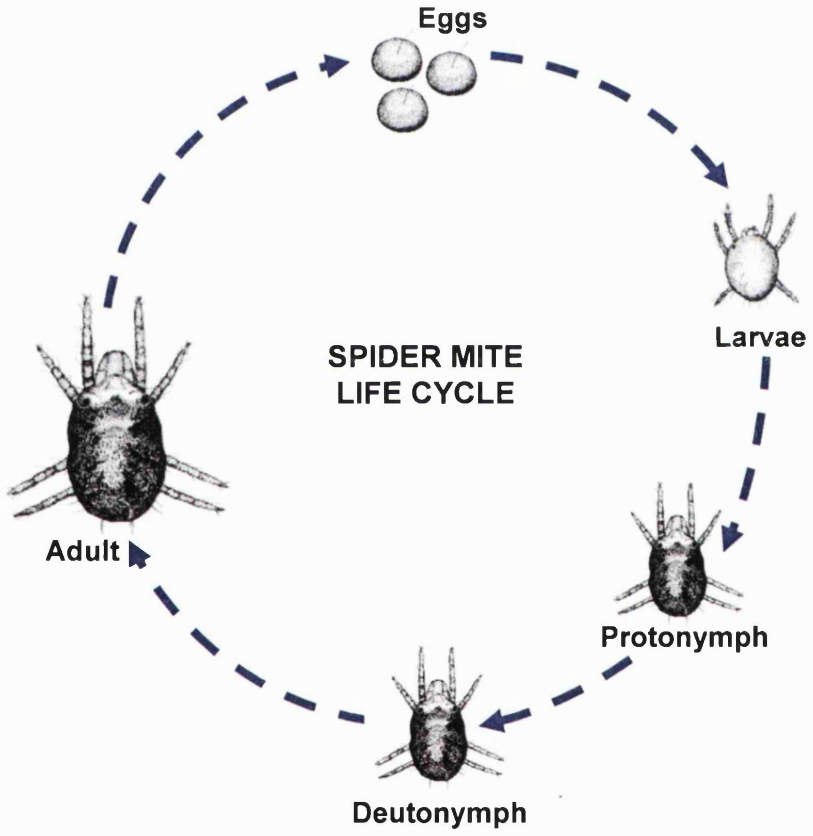



**Figure 3** Spider mite life cycle consists of four active stages.

4.1.5 Economic importance

Spider mites are major economic pests for the agriculture industry and the common gardener. They damage plants by penetrating the plant tissue with their mouthparts and sucking the cell contents from leaf tissue. Each species has its own unique way of damaging its host plant. For example, as *T. urticae* feeds, it removes sap, resulting in the collapse of mesophyll tissue and the formation of small chlorotic spots at each feeding site. Continued feeding causes a stippled-bleached effect and later, the leaves turn yellow, grey or bronze. Complete defoliation may occur if the mites are not controlled. The premature defoliation of leaves on fruit trees, particularly during the mid to late season, does not typically affect the fruit crop of the same season, but may instead cause damage to the following year's crop. On annual vegetable plants, the loss of leaves to spider mite infection can have a serious impact on production (Burnip *et al.*, 2006; Kane and Ochoa, 2006).

The introduction of pesticides has led to an expansion of spider mite populations both in the agricultural industry and in the common garden due to a reduction in populations of their natural enemies and an increase in pesticide resistance. As an alternative and more environmentally friendly method to pest control, Phytoseiid mites, predators of spider mites, can provide a consistent biological control against mites, such as the European red mite (*Panonychus ulmi*) and the twospotted spider mite (*T. urticae*) (Croft and Luh, 2004; Prischmann *et al.*, 2002).

4.1.6 Importance of spider mites as a host study organism

Very little is known about the molecular phylogeny of *Wolbachia* spp. within spider mite populations, with most research focusing on the following two issues: i) the complex host/parasite interactions, and ii) the control of spider mite populations. Spider mites have several advantages for empirical studies of *Wolbachia* population genetics. Firstly, they have short generation intervals (development of eggs to adults can take between 2 – 3 weeks) therefore allowing the study of multiple generations within a summer season. Secondly, spider mites have a small body size and no wings, resulting in limited mobility, therefore making them ideal study organisms in the field. Thirdly, due to their small size laboratory stocks can be easily reared and

maintained at relatively low costs. Finally, the role *Wolbachia* plays in spider mite reproduction can be easily exploited and better understood within laboratory stocks.

## 4.2 METHODS

### 4.2.1 Sampling

Sampling was conducted at eleven sites across continental Europe between April and May 2004 (Fig. 4; Table 1) by Vera Ros and Dr. Hans Breeuwer (University of Amsterdam, Netherlands). Twenty female spider mites across four host species were isolated; *Bryobia kissophila* (n=17), *B. praetiosa* (n=1), *B. sarothamni* (n=1) and *Tetranychus urticae* (n=1). The host plants on which spider mites were found are English ivy (*Hedera helix*, n=17), Broad bean plant (*Vicia* sp., n=1), Scotch broom (*Cytisus scoparius*, n=1) and Rosemary/Lavender (*Rosmarinus/Lavandula*, n=1).

**Figure 4** Sampling locations across Europe. Each location is represented by a circle, colour coded according to country of origin; red – Spain; blue – France; pink – Belgium; orange – Netherlands.

| Location | Isolation site |
|---|---|
| Belgium | Lompret |
| | Vierves sur Visoin |
| France | Angers |
| Netherlands | Amsterdam |
| | Hardegarijp |
| | Valkenburg |
| | Varsseveld |
| Spain | Alcossebre |
| | Begues |
| | Gatova |
| | Valencia |

**Table 1** Location of spider mite sampling.

4.2.2 Isolation of samples

Samples were recovered as described previously in Chapter 2, Section *2.2 Preparation and storage of cell and DNA stocks.*

4.2.3 Isolate nomenclature

As described in Chapter 3 Section 2.3 *Isolate nomenclature.*

4.2.4 Choice of genes, primer design & PCR conditions

As described in Chapter 3 Section 2.4 *Choice of genes, primer design & PCR conditions.*

4.2.5 Nucleotide sequence analysis

-Sequence editing and alignment

As described in Chapter 2 Section 2.5 *Nucleotide sequence analysis.*

4.2.6 Phylogenetic analysis

-Distance methods

As described in Chapter 2 Section 2.6 *Phylogenetic Analysis.*

-Bayesian inference of phylogeny

As described in Chapter 2 Section 2.6 *Phylogenetic Analysis* and Chapter 3 Section 2.6 *Phylogenetic analysis*

4.2.7 Tests for recombination

-Tests of neutrality

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Sawyer's Run Test

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Population-scaled recombination rate (ρ)

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-DnaSp

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Splits decomposition

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

4.2.8 Other methods of analysis

-eBURST

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

-Datamonkey

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

## 4.3 RESULTS

### 4.3.1 The dataset

A total of 20 strains were isolated and sequenced at 3 housekeeping loci; *gltA, groEL* and *trmD*. In addition, *wsp*, the gene encoding the *Wolbachia* surface protein and *ftsZ*, the bacterial cell-cycle gene, were sequenced for each isolate. For details of eukaryotic host and geographical location, please see Appendix B.

### 4.3.2 Sequence parameters

The alleles defined for the study were between 366 bp (*gltA*) and 681 bp (*ftsZ*) in length, and between 3 (*gltA*) and 10 (*trmD*) alleles were present per locus (Table 2). In order to gauge the selective pressure of the protein-coding genes, the dS/dN ratio was calculated in addition to the average pairwise divergence, $\pi$. The most evolutionary conserved protein-coding gene is *ftsZ*, with the highest dS/dN ratio (7.67) and the median average pairwise divergence ($\pi$ = 0.008). In addition, this locus has low levels of polymorphisms (3.2 %) and a total of 6 alleles. The second most conserved locus, with a relatively low dS/dN ratio (3.22) is *trmD*. With the highest average pairwise divergence ($\pi$ = 0.014) and percentage of variable sites (5 %) of the MLSA loci, and the greatest number of alleles (10) across the five loci, *trmD* is under moderate levels of diversifying selection. *gltA* has a dS/dN ratio of $\infty$, which is suggestive of positive selection, however, this locus has only two variable sites (0.5 % variable sites), both of which are nonsynonymous thus biasing the interpretation of the dS/dN ratio. This locus also has the lowest number of alleles (3), level of $\pi$ (0.002), and by using the selection detection programme Datamonkey, no positively or negatively selected sites were observed (data not shown). Therefore, *gltA* is under purifying selection. *groEL* has the lowest dS/dN ratio (0.5) of the five loci, suggesting it has the greatest level of positive selection. However, *wsp* with a dS/dN ratio of 1.05 has a significantly higher level of $\pi$ (0.005 & 0.062, respectively) and percentage of variable sites (1.8 % & 21.3 %, respectively) than *groEL* ($P < 0.0001$). This is expected as *wsp* encodes *Wolbachia* surface protein and contains a number of hypervariable regions.

| Loci | Fragment size (bp) | No. alleles | % Variable sites | dS/dN | $\pi$ |
|---|---|---|---|---|---|
| *gltA* | 366 | 3 | 0.5 | $\infty$ | 0.002 |
| *groEL* | 492 | 6 | 1.8 | 0.5 | 0.005 |
| *trmD* | 456 | 10 | 5 | 3.22 | 0.014 |
| *ftsZ* | 681 | 6 | 3.2 | 7.67 | 0.008 |
| *wsp* | 615 | 7 | 21.3 | 1.05 | 0.062 |

**Table 2** Genetic diversity of *Wolbachia* spp. loci.

Watterson's population mutation rate ($\theta$) and Hudson's R parameter were calculated for each of the MLSA loci, *wsp* and *ftsZ* (Table 3). Of the tree MLSA genes, *gltA* and *groEL* have the lowest population mutations rates, 0.716 and 2.537 respectively, however *groEL* has the lower population recombination rate of the two loci (0.001) and therefore is the more evolutionary conserved of the genes. *gltA* has the apparently the highest recombination rate but this is likely to be an artefact arising from the very small number of polymorphic sites. Due to only two nonsynonymous recombination events within the small region sequenced (366 bp), the recombination rate is distorted and does not accurately reflect the true rate of recombination within the population. *trmD* has the highest population mutation rate (6.589) and recombination rate (3.9) of all five loci. This is surprising as *trmD*, which encodes a tRNA methylase, has a higher population recombination rate than that of *wsp* (2.5), which has a substantially greater population mutation rate (31.395), the greatest of all the loci. *ftsZ* has a similar, but marginally lower population mutation (5.574) and recombination rates (3.1) to *trmD*.

| Loci | θ | R |
|---|---|---|
| *gltA* | 0.716 | > 1000 |
| *groEL* | 2.537 | 0.001 |
| *trmD* | 6.589 | 3.9 |
| *ftsZ* | 5.574 | 3.1 |
| *wsp* | 31.395 | 2.5 |

**Table 3** Population mutation rate (θ) & recombination rate (R) of *Wolbachia* spp. (n=20).

Tajima's D and Fu and Li's D* & F* tests were calculated to assess the level of neutrality within the six loci (Table 4). No evidence of selection was observed for the five loci, with each locus failing to provide significant evidence of selection in each of the three neutrality tests ($P > 0.01$). The three house keeping genes (*gltA, groEL* and *trmD*) were positive for Tajima's D test suggesting an excess of intermediate-frequency variants, with *ftsZ* and *wsp* producing negative values; none of these observations were significant ($P > 0.10$). Two loci were positive for Fu and Li's F* and D* test (*gltA & groEL*), although these were not significant. The three negative loci for the two tests (*trmD, ftsZ* and *wsp*) were also not significant ($P > 0.10$), and may suggest an excess of rare polymorphisms within these genes due to a recent population expansion or selective sweep. *trmD* appears to have been less affected by any potential selective sweep than *wsp* and *ftsZ* as its Fu and Li values are closer to zero.

| Loci | Tajima's D | Fu & Li's F* | Fu & Li's D* |
|---|---|---|---|
| *gltA* | 0.63240 (P > 0.10) | 0.92146 (P > 0.10) | 0.86615 (P > 0.10) |
| *groEL* | 0.00002 (P > 0.10) | 0.28605 (P > 0.10) | 0.34610 (P > 0.10) |
| *trmD* | 0.06315 (P > 0.10) | -0.01232 (P > 0.10) | -0.04147 (P > 0.10) |
| *ftsZ* | -0.38767 (P > 0.10) | -1.23407 (P > 0.10) | -1.34272 (P > 0.10) |
| *wsp* | -0.23122 (P > 0.10) | -1.07858 (P > 0.10) | -1.22445 (P > 0.10) |

**Table 4** Tajima's D, Fu & Li's F* and Fu & Li's D* test of *Wolbachia* spp. (n=20).

4.3.3 Phylogenetic analysis

Bayesian trees were constructed for each locus as described in the Methods (Fig. 5a-e) and for the concatenated MLSA sequences (Fig. 5f). The topologies of *ftsZ* and *wsp* are largely congruent, with five times more variation observed within *wsp* than *ftsZ*. As *ftsZ* is the more conserved of the two genes, it has been used as the standard to which other genes are compared. *ftsZ* phylogeny reveals two major groups (Fig. 5d). The group highlighted in red and located at a basal position in Figure 5d consists of six isolates. The second group, highlighted in blue is larger and consists of twelve isolates and is well differentiated from the smaller basal cluster (posterior probability = 98 %). Intermediate to these two groups lies two isolates, the position of both of these isolates is well supported (posterior probability > 97 %) and they are highlighted on Figure 5d in green.

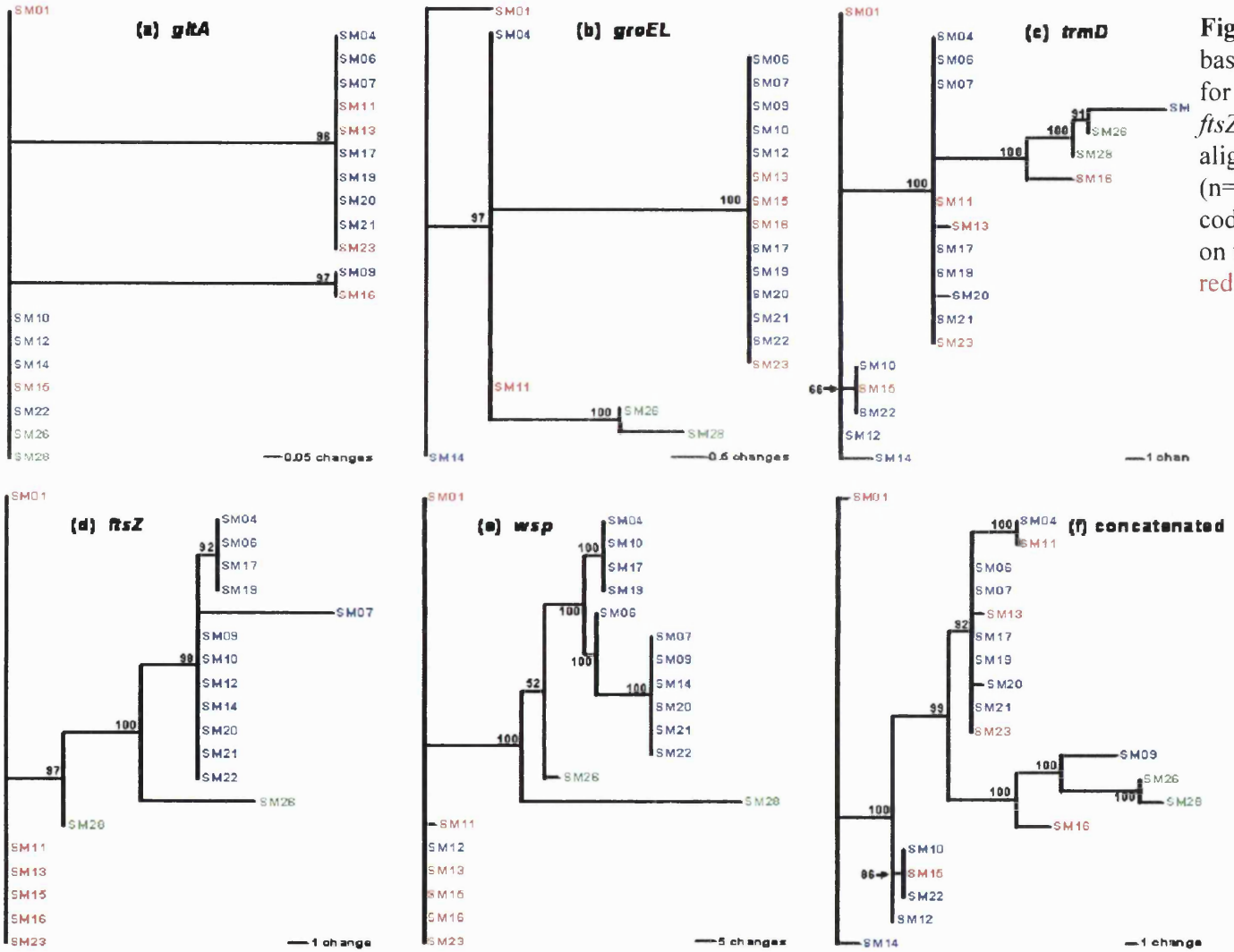



**Figure 5** Phylogenetic trees based upon the Bayesian method for each of the 3 MLSA loci, *ftsZ*, *wsp* and the concatenated alignment of the MLSA loci (n=20). Each strain is colour coded according to its position on the phylogenetic tree, i.e., red, green or blue.

In order to visually compare the topologies of the other gene trees, all strains were colour coded according to their clustering at *ftsZ*. The red group was consistent between *ftsZ* and *wsp*, however with the housekeeping loci large inconsistencies were observed. Both *gltA* and *groEL* (Fig. 5a & 5b) are the most conserved of the five loci with relatively little variation observed (scale: 0.05 changes) and as such are too uniform to interpret this pattern of clustering with any confidence. Isolates SM11, SM13, SM16 and SM23 cluster with the blue clade at *gltA, groEL* and *trmD* (posterior probability > 96 %). SM15 also shows inconsistent phylogenetic position on the different gene trees and clusters with the blue clade at *groEL*, but with the red clade at all other loci. SM10, SM12, SM14 and SM22 cluster not with the blue clade on *gltA* and *trmD,* but instead with the red clade. Also SM04 and SM14 do not group with the blue clade on *groEL*. SM12 clusters with the red group at *gltA*, *trmD* and *wsp*. The intermediate isolates (highlighted green) repeatedly cluster together across the five loci. These strains are observed within the red clade on *gltA*, within the blue clade on *groEL* and *trmD,* and on *ftsZ* and *wsp* appear to be hybrids of the two main groups (Fig. 6). The concatenated tree (Fig. 5f) resolves the inconsistencies within the single housekeeping loci, however is discordant with *ftsZ* and *wsp* phylogenies. The tree does not clearly distinguish between the red and blue groups of isolates and places the intermediate green isolates within the large blue clade on the tree. The concatenation of MLSA loci, *ftsZ* and *wsp* show phylogenetic concordance between *wsp* and *ftsZ* topologies and discordance with the three MLSA loci (Fig. 7).

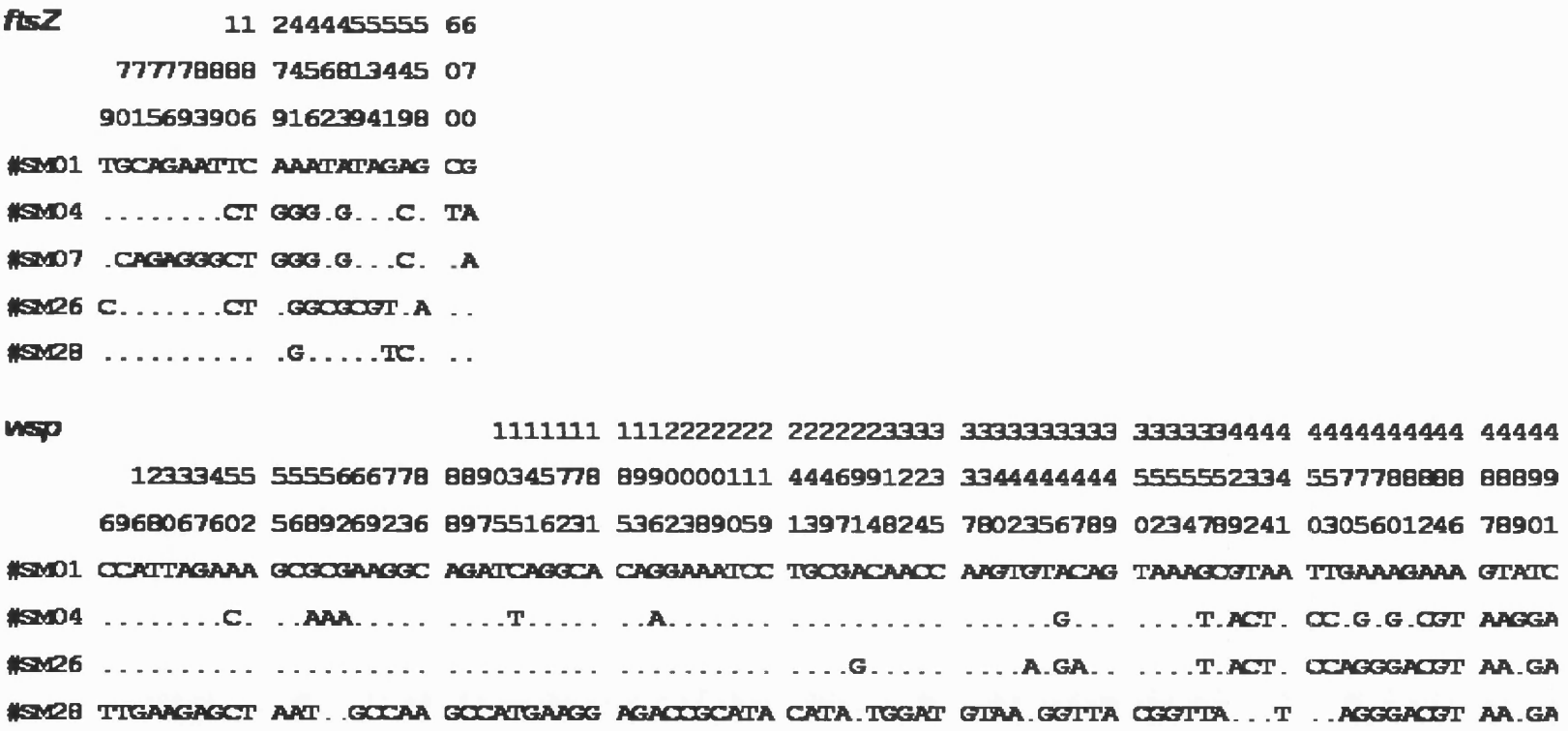



**Figure 6** Evidence of mosaicism within *ftsZ* and *wsp* (SM26 & SM28) resulting from horizontal transfer between the red group (SM01) and blue group (SM04).

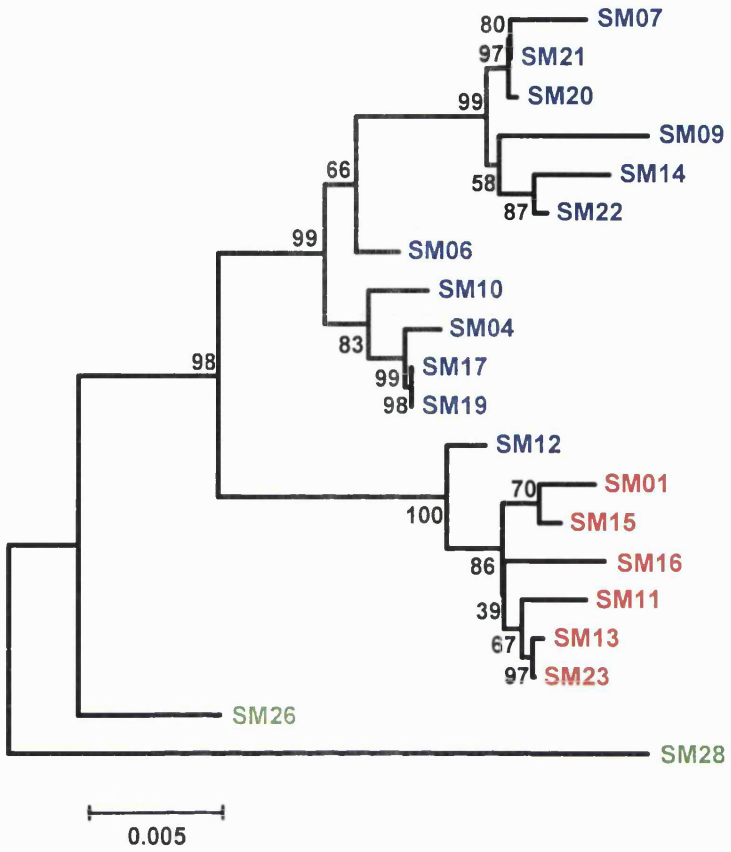



**Figure 7** Neighbour-Joining tree based upon the concatenation of MLSA loci, *ftsZ* and *wsp* (n=20). Each strain is colour coded according to its position on the phylogenetic tree according to *ftsZ*, i.e., red, green or blue.

4.3.4 Host association observed amongst isolates

Four species of spider mite host were included in the analysis; *B. kissophila* (n=17), *B. praetiosa* (n=1), *B. sarothamni* (n=1) and *T. urticae* (n=1). *ftsZ* has again been used to determine the level of host association (Fig. 8a–e). Across the 3 MLSA loci, *B. sarothamni* and *T. urticae* (SM26 and SM28 respectively) cluster together and at *wsp* and *ftsZ* they are placed between the major *B. kissophila* clades. In most cases, with the exception of *gltA*, this clustering is distinct with high posterior probability support (97 % - 100 %). *B. sarothamni* and *T. urticae* cluster with multiple *B. kissophila* isolates at *gltA* forming a clade on the backbone of the tree (Fig. 8a). *B.*

*kissophila* spans both the red and blue clusters discussed in the previous section and is divided into three clades (Fig. 8d). Clade 1 consists of SM01, SM11, SM13, SM15, SM16 and SM23; clade 2 consists of SM07, SM09, SM12, SM14, SM20, SM21 and SM22; clade 3 consists of SM04, SM06, SM17 and SM19. The single *B. praetiosa* isolate (SM16) clusters with clade 1 of the *B. kissophila*, and *B. sarothamni* and *T. urticae* are observed intermediate to *B. kissophila* clades 1 and 2. Although largely congruent, some inconsistencies are apparent between *ftsZ* and *wsp*, for example, SM10 at *wsp* clusters with clade 3 instead of clade 2 at *ftsZ*, SM12 clusters with clade 1 instead of clade 2 and SM06 clusters with clade 2 instead of clade 3 (Fig. 8e). These inconsistencies are a result of horizontal transfer of *Wolbachia* within *B. kissophila* (Fig. 9), with localised replacements resulting in mosaic structure observed in isolates SM12 and SM06 (data not shown). Within *trmD*, little congruence is observed between the three *B. kissophila* clades (Fig. 8c). SM09 no longer groups with clade 2 of *B. kissophila* and instead clusters with *B. praetiosa, B. sarothamni* and *T. urticae*; this is as a result of horizontal transfer (Fig. 10).

The high level of congruence between *ftsZ* and *wsp* phylogenies and their distant positions on the *Wolbachia* genome suggests evidence of host association. This is surprising due to clear evidence of mosaicism observed within the two loci (Fig. 6 & 9) and suggests that recombination has not been so pervasive as to completely erode any signal of host /*Wolbachia* co-evolution. To determine the level of congruence between *Wolbachia* and host phylogeny, the mitochondrial gene encoding cytochrome oxidase 1 (*CO1*) was sequenced for each host by Vera Ros of the University of Amsterdam (data not shown). Unfortunately these data did not contain sufficient resolution to infer the relevance of the three distinct *B. kissophila Wolbachia* clades evident from *wsp* and *ftsZ*.

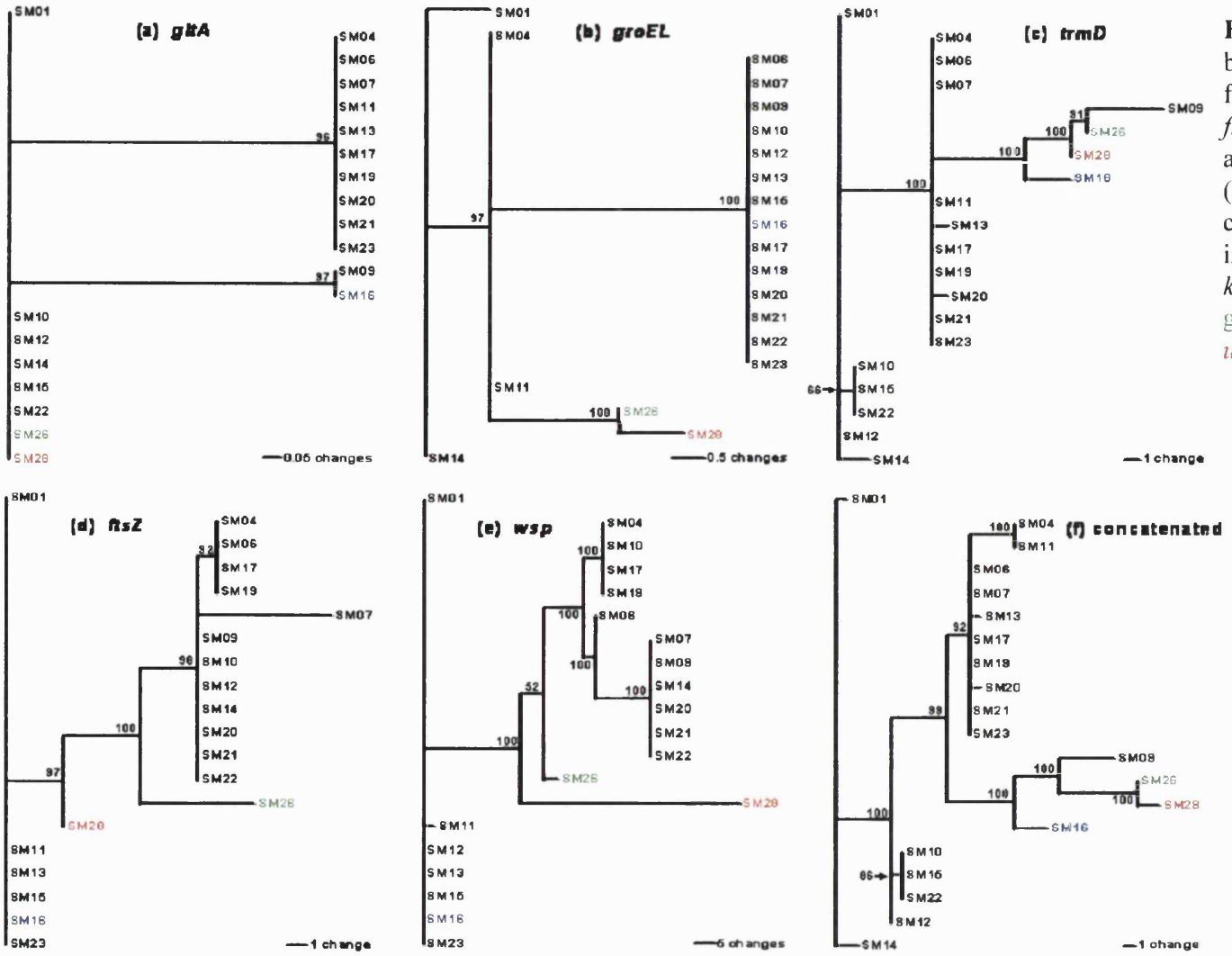



**Figure 8** Phylogenetic trees based upon the Bayesian method for each of the 3 MLSA loci, *ftsZ*, *wsp* and the concatenated alignment of the MLSA loci (n=20). Each strain is colour coded according to its invertebrate host; black = *B. kissophila*; blue = *B. praetiosa*; green = *B. sarothamni*; red = *T. urticae*.

```
          111111 1112222222 2333444444 4444444444 4445555555 555
       5556378899 9990000112 3445233557 8888888899 9990000000 111
       0892521434 6790124293 1677924035 0245678901 2490123467 034
#SM01  AGCGCGAGAA GCTCTGATCT AACGGTATTA AAACAGTATC ACAAAGAGTC CCC
#SM10  CAAAT..... A......... .G.TACTCCG GCG.TAAGGA TAGGTATCAA TTG
#SM09  CA...AGATC .AAACAGCTG GGATACTCCG GC.ATAAGGA TAGGTATCAA TT.
```

**Figure 9** Evidence of mosaicism within *wsp* (SM10) resulting from horizontal transfer between the clade 1 of *B. kissophila* (SM01) and clade 2 of *B. kissophila* (SM09).

```
                11222 2334444444
          1124728011 8480123334
          2922847407 8879283582
#SM01 CGAACAGCAT AACGGGATTA
#SM10 .......... .......G..
#SM04 .....GA..C G.T.....C.
#SM09 TAGTTGAAG. GGTTCTG..T
```

**Figure 10** Evidence of high divergence within trmD (SM09) resulting from whole-allele horizontal transfer.

4.3.5 Phylogeny of *Wolbachia* spider mite supergroups

*ftsZ* sequence data was aligned with sequences from previous *Wolbachia* supergroup studies, to determine the *Wolbachia* supergroup phylogeny of spider mites (Baldo *et al.*, 2005a; Casiraghi *et al.*, 2005; Panaram and Marshall, 2006; Woo Oh *et al.*, 2000). Sequences representing *Wolbachia* supergroups, A – F and H, were retrieved from GenBank (Table 5). No alignments were made for *Wolbachia* supergroup G due to lack of available *ftsZ* data.

| Host species name | Accession Number | Supergroups |
|---|---|---|
| *Camponotus vafer* | DQ266396 | A |
| *Protocalliphora* sp. | DQ266422 | A |
| *Nephila clavata* | AF232235 | A |
| *Acraea eponina* | DQ266520 | B |
| *Nasonia giraulti* | DQ266527 | B |
| *Teleogryllus taiwanemma* | DQ266529 | B |
| *Dirofilaria immitis* | AJ010272 | C |
| *Wuchereria bancrofti* | AF081198 | D |
| *Folsomia candida* | AJ344216 | E |
| *Kalotermes flavicollis* | AJ292345 | F |
| *Zootermopsis angusticollis* | AY764283 | H |

**Table 5** Host species and accession numbers representing each of the *Wolbachia* supergroups, with the exception of supergroup G (Baldo *et al.*, 2005; Casiraghi *et al.*, 2005; Panaram and Marshall, 2006; Woo Oh *et al.*, 2000).

As discussed in Section 4.1.2, earlier supergroup studies have placed spider mites in *Wolbachia* supergroup B (Breeuwer and Jacobs, 1996). Our data corroborates previous findings with all twenty *Wolbachia* spider mite isolates clustering within supergroup B (Fig. 11). The spider mite samples are divided into two clusters within this group. The cluster located at the top of the tree consists of isolates from *B. kissophila* clades 2 & 3 (n=12) and *B. sarothamni* (n=1). The second cluster consists of isolates from *B. kissophila* clade 1 (n=6) and *T. urticae* (n=1). Baldo *et al.* recently reported a subdivision of *Wolbachia* supergroup B observed repeatedly within a number of loci (Baldo *et al.*, 2005). Figure 11 strongly supports a bifurcation of supergroup B into two distinct clades i) *Encarsia formosa* & *Protocalliphora sialia* (bootstrap support = 100 %) and ii) the remaining spider mite and supergroup B isolates.

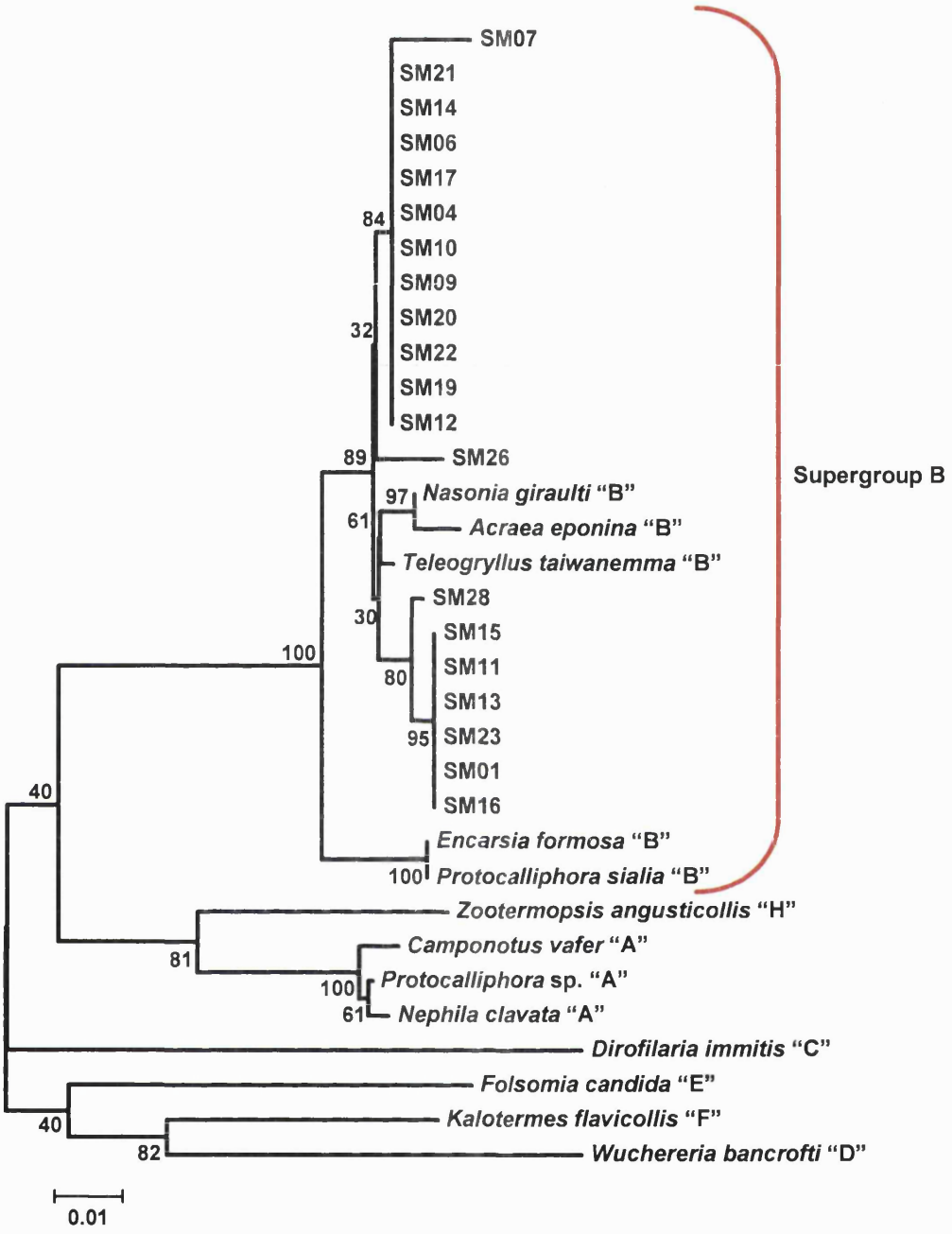



**Figure 11** Neighbour-Joining tree for *ftsZ* in which *Wolbachia* spider mite isolates cluster with *Wolbachia* supergroup B (n=20). Spider mite data was aligned with sequences retrieved from Genbank from previous supergroup studies. In inverted commas is the *Wolbachia* supergroup that particular species belongs to.

4.3.6 Biogeographical adaptation within *Wolbachia* spp.

Characterizing geographic patterns of genetic variation within and among populations is a necessary precursor to understanding mechanisms of population differentiation and speciation. In Chapter 3 global geographical structuring of *Wolbachia* spp. was not apparent between the numerous species of arthropods in the MLSA dataset, however this analysis did not address the extent of fine scale geographical structure between *Wolbachia* spp. associated with closely related hosts.

To address the extent of geographical structuring of *Wolbachia* spp., the five loci were colour coded according to geographical source (Fig. 12a-e). Across each loci, there is little obvious geographical pattern, with spider mites isolated from a single source often clustering within more than one clade. Local geographic structure is limited however some loci show clear examples of geographic structuring. For example, in *ftsZ* clustering is observed between isolates from Amsterdam (SM04, SM17) and Hardegarijp (SM06, SM19), however not with isolates from Valkenburg (SM09) and Varsseveld (SM21) in the Netherlands. The incongruence amongst the phylogenetic trees indicates *Wolbachia* has undergone numerous recombination events throughout its evolutionary history. *ftsZ* and *wsp* (Fig. 12d & e), the two loci at which the greatest degree of host association is observed, lack clear geographical structure within the spider mite population. Instead structure is observed within *trmD* between "Northern", i.e. Dutch, French and Belgian isolates, and "Southern", i.e. Spanish isolates (Fig. 12c). A geographical boundary such as the Pyrenees might explain such structuring. One Spanish spider mite strain is found in the Northern grouping, however this is likely due to this *Wolbachia* strain (SM28) having been isolated from a different spider mite species (*T. urticae*) which has undergone horizontal transfer (Fig. 13). *gltA,* to a lesser extent, also shows the Northern and Southern populations, with only one isolate (SM26, a Northern strain which has gained a Southern *gltA* locus) conflicting to this rule. The difference between the Northern and Southern populations in *gltA* is based on only a single nucleotide change therefore should be treated with caution.

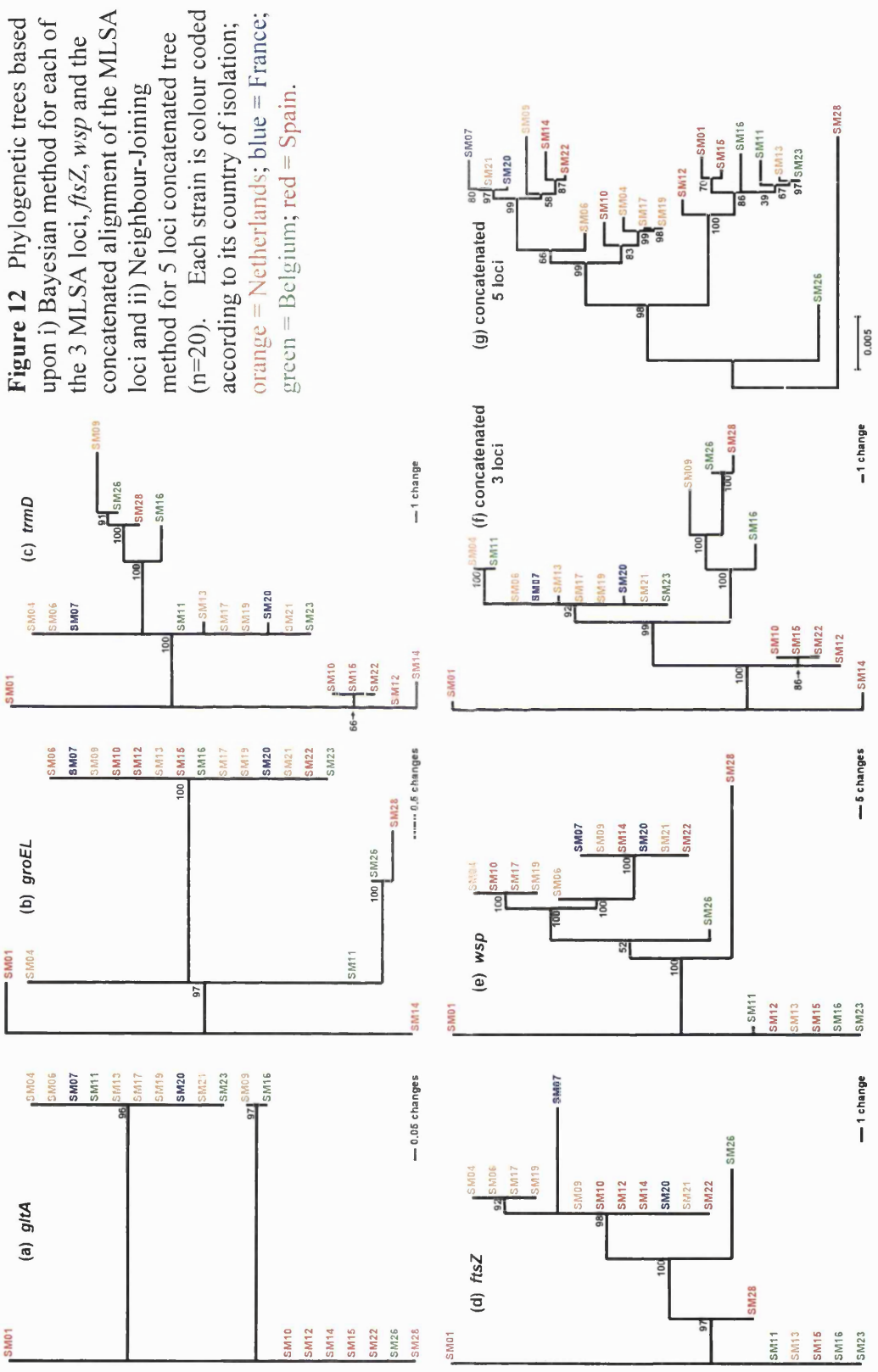



**Figure 12** Phylogenetic trees based upon i) Bayesian method for each of the 3 MLSA loci, *ftsZ*, *wsp* and the concatenated alignment of the MLSA loci and ii) Neighbour-Joining method for 5 loci concatenated tree (n=20). Each strain is colour coded according to its country of isolation; orange = Netherlands; blue = France; green = Belgium; red = Spain.

```
                  11122 2344
             1124725801 8833
             2922846740 8735
#SM01        CGAACATGCA ACAT
#SM12        .......... ....
#SM14        ......C... ..G.
#SM15        .......... ...G
#SM10        .......... ...G
#SM22        .......... ...G
#SM28        TAGTTG.AAG GT..
```

**Figure 13** Evidence of mosaicism within *trmD* Spanish spider mite isolates SM28.

To address the extent of gene flow and genetic differentiation between the proposed Northern and Southern populations, Wrights inbreeding coefficient ($F_{ST}$) was calculated (Table 6). The $F_{ST}$ values for *groEL, ftsZ* and *wsp* are all negative indicating these loci are undergoing higher rates of gene flow. Consequently, there is relatively little evidence for geographic structure within these loci. In contrast, both *trmD* and *gltA* have moderately high $F_{ST}$ values (0.62514 and 0.58974 respectively) indicating non-random distribution of genotypes between north and south Europe and limited gene flow.

| Loci | $F_{ST}$ value |
|---|---|
| *gltA* | 0.58974 |
| *groEL* | -0.02457 |
| *trmD* | 0.62514 |
| *ftsZ* | -0.08056 |
| *wsp* | -0.07206 |

**Table 6** Wrights inbreeding coefficient ($F_{ST}$), comparing Northern (n=14) and Southern (n=6) *Wolbachia* spp.

To determine the overall phylogenetic picture, each of the loci were concatenated to produce i) MLSA 3 loci and ii) 5 loci concatenated trees (Fig. 12 f & g); these were colour coded accordingly. The level of geographic structure within Figure 12f is similar to that of *trmD* due to the conserved nature of *gltA* and *groEL* which make little contribution to the concatenated dataset. When sequence information for *ftsZ* and *wsp* is added to the concatenated data the geographic structure between the Northern and Southern isolates is lost, which is consistent with the $F_{ST}$ values showing little evidence of a North/South division for these genes. Curiously, therefore, the data are supportive of two different phylogenies. On the one hand, *wsp* and *ftsZ* divide the *B. kissophila* isolates into three clades, which may possibly reflect cryptic co-evolution with the host but do not correspond to geographical source. On the other hand, the signal from *gltA*, *groEL* and particularly *trmD* do not support these three groups, but instead provide evidence for geographical structuring, and in particular a distinction between isolates sampled from the North and South of Europe.

### 4.3.7 Genetic variation between *Wolbachia* spp isolated from sexual and asexual hosts

As a reproductive manipulator of arthropods, most *Wolbachia* studies have focused on the cause and effect of these manipulations on the host rather than examining the level of genetic variation within sexual/asexual *Wolbachia* spp. To address this, the five loci were colour coded according to the type of reproduction observed (Fig. 14a-e), i.e. sexual (n=2) and asexual (n=18). At all loci, the sexual *Wolbachia* spider mites are observed grouping together, however not in all cases forming a distinct group on the phylogenetic tree. *ftsZ* and *wsp* (Fig. 14 d & e) show the sexual spider mites located centrally on the tree, between clade 1 and 2 of *B. kissophila.* Within *groEL* (Fig. 14b), one of the more conserved housekeeping genes, both isolates group together (posterior probability = 100%) forming a distinct lineage from the remaining asexual spider mites. Both *gltA* and *trmD* show the sexual spider mites clustering with asexual isolates. At *trmD* isolates SM09 and SM16, both asexual strains, form a discrete cluster with the two sexual spider mite isolates and have undergone recombinational exchange (Fig. 15). Within *gltA* the sexual spider mites are identical to asexual "Southern" isolates. It would be expected that sexual spider

mites have encountered more genetic variation and would therefore form discrete lineages from asexual spider mites. Although the small number of isolates from sexual species means it is not possible to examine this, the horizontal exchange between the sexual and asexual mites is surprising and implicates mechanisms for transfer between hosts other than host sexual reproduction (see Chapter 3).

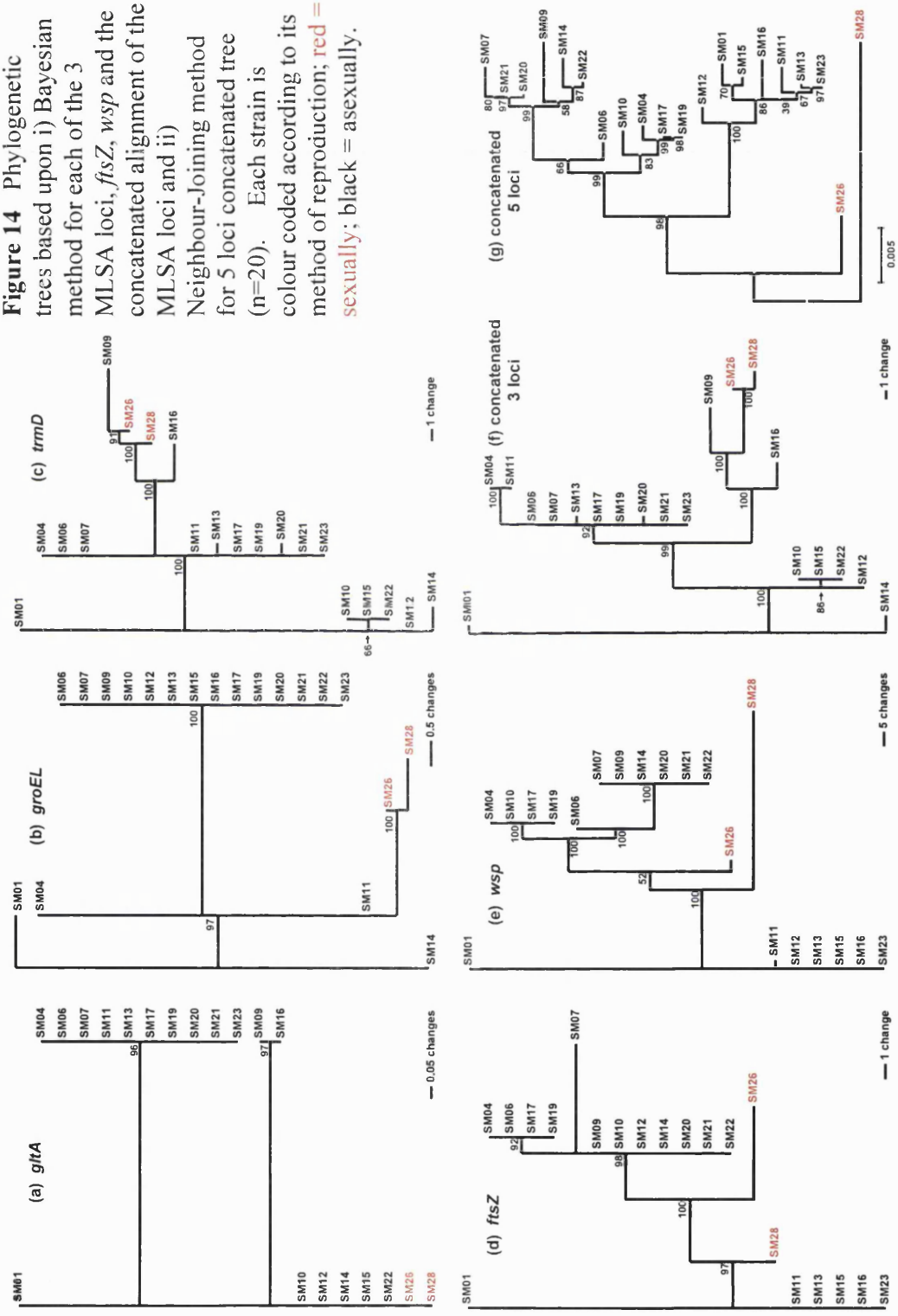



**Figure 14** Phylogenetic trees based upon i) Bayesian method for each of the 3 MLSA loci, *ftsZ*, *wsp* and the concatenated alignment of the MLSA loci and ii) Neighbour-Joining method for 5 loci concatenated tree (n=20). Each strain is colour coded according to its method of reproduction; red = sexually; black = asexually.

```
                  112 2223344444 44
           1124667280 1184801233 34
           2922138474 0788792835 82
#SM01 CGAAGGCAGC ATAACGGGAT TA
#SM04 .......GA. .CG.T..... C.
#SM16 TAGTAATGA. .CG.T....G ..
#SM26 TAGT..TGAA G.GGT..... ..
#SM28 TAGT..TGAA G.G.T..... ..
#SM09 TAGT..TGAA G.GGTTCTG. .T
```

**Figure 15** Evidence of mosaicism within *trmD* within sexual (SM26 & SM28) and asexual (SM09 & SM16) spider mite populations.

4.3.8 Evidence for recombination within the *Wolbachia* spider mite population

To examine the extent of recombination within the population, the coalescence approach of Fearnhead & Donnelly was performed for each of the five loci (Fearnhead and Donnelly, 2001). Three loci, *gltA*, *ftsZ* and *wsp*, show the minimum value for ρ suggesting these loci are not subject to recombination (Table 7). This is surprising as high levels of recombination within *wsp* would be expected due to its function as an outer surface protein. *groEL* has a ρ value of 1.01 demonstrating this loci is under low level selective pressures. The highest level of recombination within the population is observed with *trmD* (7.071), consistent with the high level of mosaicism within this gene. The mean ρ value for the MLSA loci (ρ = 2.69) is higher than that observed in Chapter 3, although is still comparable with human pathogenic bacteria (Perez-Losada *et al.*, 2006).

| Loci | LDhat estimate of ρ |
|---|---|
| *gltA* | 0 |
| *groEL* | 1.01 |
| *trmD* | 7.071 |
| *ftsZ* | 0 |
| *wsp* | 0 |

**Table 7** Estimations of the recombinational parameter ρ within each MLSA loci, *ftsZ* and *wsp* using LDhat (n=20).

The Sawyer's Runs Tests were also used to determine the extent of recombination within the *Wolbachia* dataset (Sawyer, 1989). This test looks for evidence of recombinational exchange within a set of aligned sequences (Table 8).

| | Method of analysis | |
|---|---|---|
| Loci | SSCF (P-value) | SSUF (P-value) |
| *gltA* | 0 (P > 0.05) | 1 (P > 0.05) |
| *groEL* | 0 (P > 0.05) | 1 (P > 0.05) |
| *trmD* | 1324 (P > 0.05) | 0.3909 (P > 0.05) |
| *ftsZ* | 651 (P > 0.05) | 4120005 (P > 0.05) |
| *wsp* | 796 (P < 0.05) | 2864680 (P > 0.05) |

**Table 8** Summary of recombinational exchange using Sawyer's Runs Test.

Neither the SSCF nor SSUF statistics revealed significant "mosaic" structure (a non-random distribution of polymorphisms) within any of the five loci with the exception of *wsp* ($P < 0.05$) for SSCF. Sawyer's Runs Test examines the rate at which recombination has effected the distribution of polymorphisms within aligned sequences, it is therefore insensitive to recombination events which have resulted in

the replacement of whole alleles. As an outer surface protein *wsp* is subject to positive selection and therefore higher levels of recombinational exchange. The four hypervariable regions in *wsp*, interspersed by conserved regions are reflected in the significant SSCF value. As discussed in Chapter 3, the remaining loci may have generally undergone large-scale replacements of the entire alleles. Such wholesale replacement of alleles is evident from the phylogenetic inconsistencies noted between the housekeeping genes and *ftsZ* and *wsp* (as discussed above).

The Homoplasy Test (Maynard Smith and Smith, 1998), which analyses the true homoplasies among informative, synonymous polymorphic sites, was used to determine the extent of recombination amongst the *Wolbachia* MLSA data. The test could only be performed with *wsp* as the housekeeping loci and *ftsZ* all has less than the minimum number of informative sites needed for the analysis. The Homoplasy ratio obtained for *wsp* (-0.107) was not significant ($P > 0.05$), indicating significantly fewer homoplasies than would be expected under free recombination.

To assess the degree of linkage disequilibrium between alleles, the Index of Association ($I_A$) was performed (Smith *et al.*, 1993). The $I_A$ for the MLSA loci was found to be 0.397 and 0.504 when all five loci are included. This indicates a moderate rate of recombination resulting in the incomplete loss of linkage between alleles.

### 4.3.9 Splits decomposition analysis

As a further test for recombination, splits decomposition analysis was performed (Bandelt and Dress, 1992). Similar phylogenetic clustering to the traditional phylogenetic approaches was observed (Fig. 16); the splits decomposition networks divided the strains either according to host association or to geographical association. The splittability index, the goodness of fit to the data, is presented in Table 9; these estimates range from 76.92 % (*trmD*) to 100 % (*gltA, groEL, ftsZ*). Where no reticulations are present, the graph presents as a bifurcating tree. Conversely, reticulations in splits graphs are indicators of conflicts in the data caused by recombination. No reticulations were present at *gltA*, *groEL*, *trmD* and *ftsZ*, therefore suggesting these loci have arisen by clonal diversification. *gltA* has the

simplest graph structure and is consistent with the hypothesis that the Northern population was seeded from the South, as the Southern genotype is intermediate to the two Northern genotypes (and basal on the tree). However, more data is required to test this hypothesis. The splits graph for *trmD* also shows a clear division between the Northern and Southern populations in addition to four highly divergent strains (*B. praetiosa*, *B. sarothamni*, *B. kissophila* and *T. urticae*). The lack of reticulations in both *gltA* and *trmD* indicate there has been limited recombination between Northern and Southern populations since their divergence. Although geographic structure is not evident within *ftsZ* and *wsp*, clear divisions within *B. kissophila* are apparent (as described above). Within *ftsZ* clades 2 and 3 of *B. kissophila* are more closely related than clade 1; both are separated by a number of long evolutionary branches. Isolates SM07 and SM26 are presented on individual long branches, with SM07 having diverged from clade 2 *B. kissophila* and SM26 intermediate to *B. kissophil*a clade 1 and clades 2 and 3. SM28 is presented on a short branch, showing greater similarity to *B. kissophila* clade 1. *wsp* is the only loci in which reticulate structure is present, with significant reticulations within the base of the graph rather than at the edges, indicating recombination has occurred between each of the *B. kissophila* clades and not yet within the three clades. *groEL* is the only loci which does not share similarities between either host and geographical structuring.

| Loci | Splittability Index |
|---|---|
| *gltA* | 100 % |
| *groEL* | 100 % |
| *trmD* | 76.92 % |
| *ftsZ* | 100 % |
| *wsp* | 92.49 % |

**Table 9** Splittability Index values for MLSA loci, *ftsZ* and *wsp* (n=20).

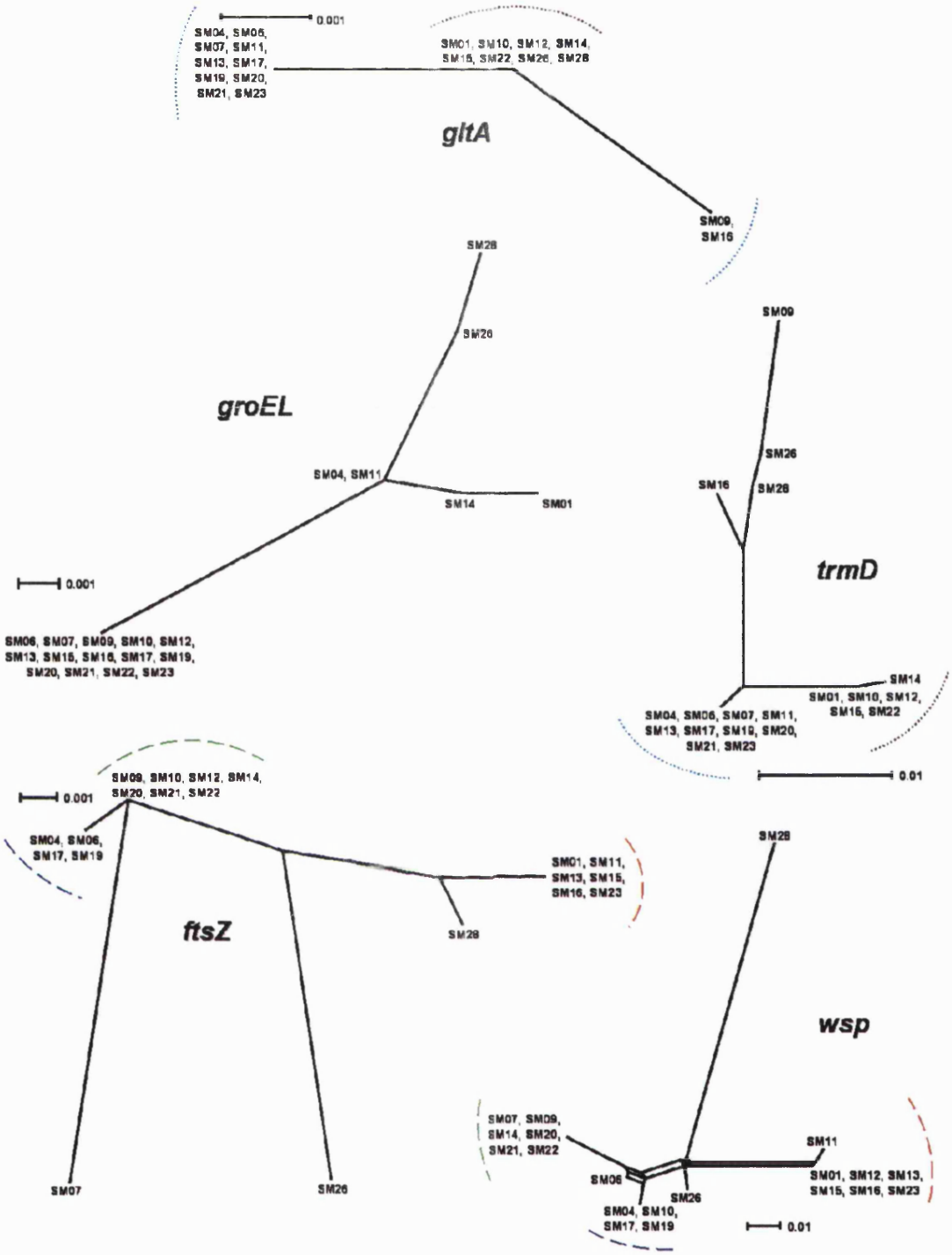



**Figure 16** Splits decomposition analysis of the MLSA loci, *ftsZ* and *wsp* (n=20). Highlighted by dotted brackets are Northern (light blue) and Southern (purple) *Wolbachia* spp. populations. Highlighted by dashed brackets are *B. kissophila* clades 1 (red), clade 2 (green) and clade 3 (blue).

### 4.3.10 Clonal diversification

Allelic profiles were assigned on the basis of the nucleotide sequence of each locus, and the structure of the population examined using eBURST (Fig. 17a). Twelve unique genotypes (STs) were determined, the majority of which were represented by only a single isolate (75 %), while three are represented by at least two isolates. The 12 STs were divided by eBURST into one minor clonal complex, one triplet and one doublet. No singletons were observed within the dataset. The largest and most common clone was ST7, which accounts for 6 isolates (30 % of all isolates) and this is the predicted founder of the minor clonal complex (bootstrap score 57 %). This clonal complex, CC7, contains 3 STs and 10 isolates, and is composed entirely of *B. kissophila* isolates. The doublet is linked to both CC7 and the triplet by DLVs. Clustering of the Southern *Wolbachia* isolates is observed within the triplet and its DLVs (ST4 and ST6) in addition to it also being composed entirely of *B. kissophila* isolates.

### 4.3.11 eBURST based on translated alleles

To adjust for the diversity of the *Wolbachia* genus, allelic profiles were re-assigned on the basis of their amino acid, rather than nucleotide sequence to view nonsynonymous evolution, and were examined using eBURST (Fig. 17b). Eleven unique aa-STs were determined, of which 54.5 % (n=6) were represented by a single isolate and 36.4 % (n=4) represented by the minimum of two isolates. In order to cross reference the ST assignments based on nt- and aa- alleles, a table is presented in Appendix B. One minor clonal complex is observed, aa-CC8, which consisting only of *B. kissophila* isolates, corresponds to nt-ST7 and the Northern spider mites. The founder of the complex, aa-ST8 is linked by DLVs to a doublet containing two different spider mite species; aa-ST9 (*B. kissophila*) and aa-ST11 (*B. praetiosa*). This doublet corresponds to the doublet observed in Figure 17a. The remaining quadruplet and associated DLVs are composed of Southern *Wolbachia* isolates. No founder ST is identified, however aa-ST2 is composed of three isolates; the remaining STs are present in only one copy. aa-ST6 corresponds to two nt-STs, nt-ST5 and nt-ST6, which are two different species of spider mite, *B. sarothamni* and *T. urticae* respectively.

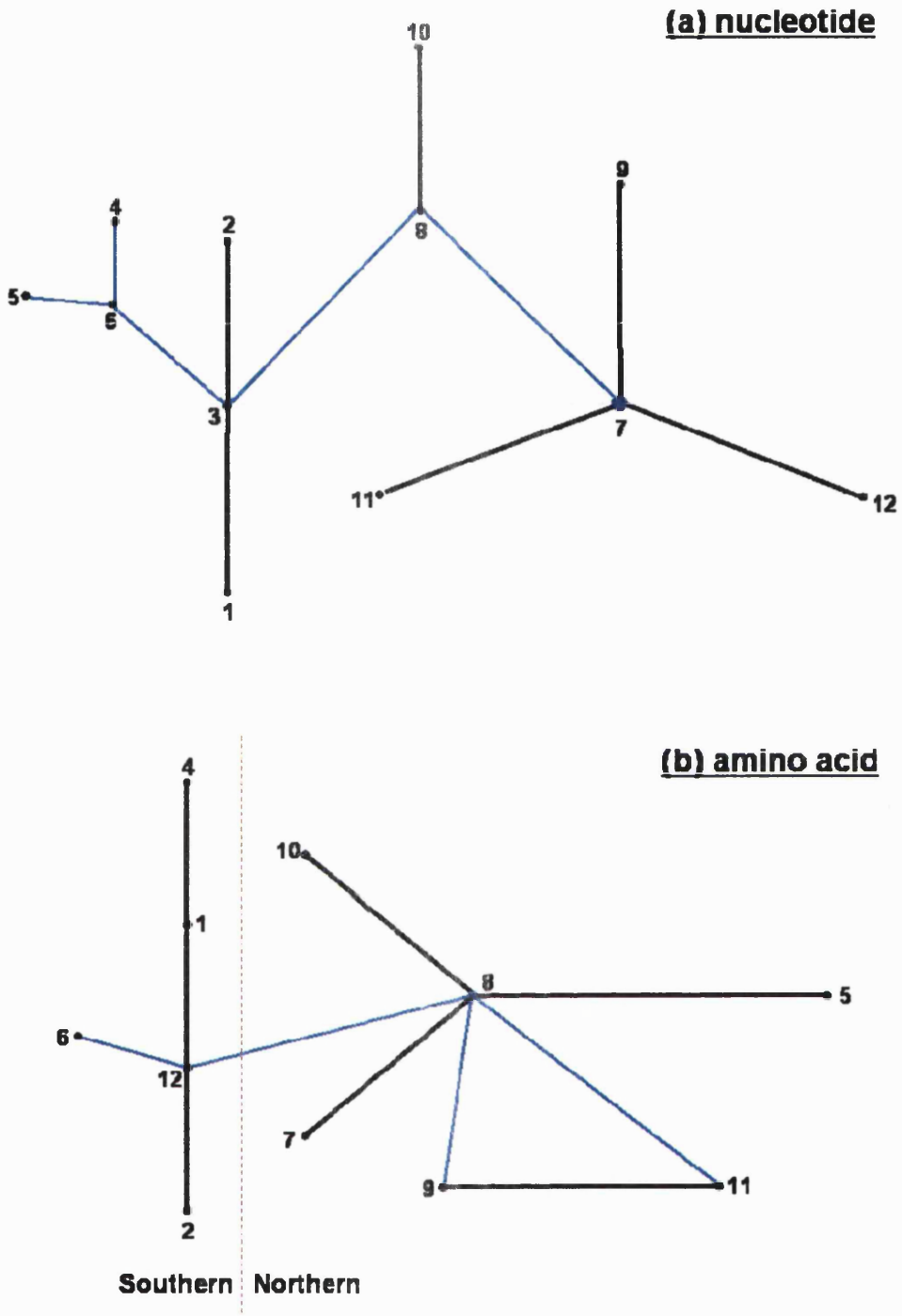



**Figure 17** A population "snapshot" showing the clusters of linked STs and unlinked STs for 20 spider mite *Wolbachia* spp. using a) nucleotide and b) amino acid sequence data (0/3 alleles in common). The predicted clonal ancestor is shown in blue, SLVs are shown as black lines and DLVs as light blue lines. The sizes of the circles that represent each ST indicate their prevalence within the population.

## 4.4 DISCUSSION

Spider mite infestation is a common problem for the agriculture industry, and as such a great deal of research has been carried out to identify suitable pest management strategies, however relatively little is known about *Wolbachia* spp. phylogeny and host specialisation within spider mites. Five loci were used in this study to examine the diversity and population structure within the Family Tetranychidae. Nucleotide sequence data for twenty isolates was generated for three MLSA loci (*gltA, groEL* and *trmD*), in addition to two traditional *Wolbachia* spp. identification markers; *ftsZ* and *wsp*. Four spider mite species (*B. kissophila*, *B. praetiosa*, *B. sarothamni* and *T. urticae*) were included in the analysis, which included a mixture of sexual (n=2) and asexual (n=18) species. These data provide the means to address the aims listed in Chapter 1 and these will now be discussed in turn:

i) Phylogenetic congruence to host phylogenies

The data presented suggest two distinct phylogenetic patterns. *gltA*, *groEL* and *trmD* provide some evidence for geographical structuring, as discussed below, whereas *ftsZ* and *wsp* are largely congruent, provide little evidence of geographical structuring and divide the *B. kissophila* isolates into three groups. The fact that these groups are consistent between these two highly variable genes supports the notion that they represent adaptive clusters, and the most likely explanation is that they represent micro-adaptation to subtle differences in the host. Although there is currently no evidence for three corresponding divisions within the *B. kissophila* population, the current data is based on mtDNA and contains no variation. mtDNA is known to be an unreliable phylogenetic marker for arthropod hosts of endosymbionts, due to the frequent occurrence of selective sweeps (Hurst and Jiggins, 2005; Rasgon *et al.*, 2006; Turelli and Hoffmann, 1991). It might be expected that host adaptation is apparent at *wsp*, given the fact that this gene encodes proteins associated with the outer membrane and is known to be immunologically relevant. Furthermore, the isolates from *B. praetiosa* and *B. sarothamni* are clearly distinct from the *B. kissophila* isolates and their relative positions on the tree are again consistent between *wsp* and *ftsZ*. This is also consistent with the suggestion of host adaptation.

This is surprising given the data presented in the previous chapter, as well as numerous other studies which have failed to find evidence for the *Wolbachia* phylogeny being parallel to that of its hosts (Jiggins *et al.*, 2001; Schilthuizen and Stouthamer, 1997; Zhou *et al.*, 1998). This challenges the notion that recombination is "pervasive" in *Wolbachia*, as recently suggested by Baldo *et al.* and that co-evolution may be occurring within the fine phylogenetic and temporal scale addressed in this study (Baldo *et al.*, 2005a). The concordance between *ftsZ* and *wsp* is striking, and may reflect a relatively low rate of recombination owing to decreased opportunity for *Wolbachia* to meet and hence recombine in the wild. Thus, the sexual abstinence of the spider mite host appears to engender a commensurate moderation in the bacterial endosymbiont.

However, it is clear that recombination does occasionally occur, which either reflects a history of sexual reproduction within the spider mite population, or an alternative means by which the bacteria are transferred between hosts in the wild. A number of mosaics are observed across both *ftsZ* and *wsp* (Fig. 6 and 9), in addition to hypervariable regions in *wsp*, indicating horizontal transmission is still taking place. Interestingly, within both *ftsZ* and *wsp* there are a number of examples of variation and mosaicism with very localised geographic scales. For example, strains SM07 and SM20 were both isolated from Angers in France, however within *ftsZ* they differ by 7 nucleotide bases within the first 13 % of the fragment sequenced (Fig. 18). These differences within populations may be the result of horizontal transfer by the mechanisms described in Chapter 3.

```
                  112 44456
          7777788887 45847
          0156939069 16390
#SM01     GCAGAATTCA AAAAG
#SM20     .......CTG GGGCA
#SM07     CAGAGGGCTG GGGCA
```

**Figure 18** Evidence of within population variation between two French spider mites strains (SM07, SM20) isolated from Angers, France.

ii) Geographical structure is observed within MLSA loci

Both *gltA* and *trmD* showed evidence of geographical structuring within the spider mite dataset. Two populations were observed; i) Southern population consisting of Spanish isolates and ii) Northern population consisting of isolates recovered from the Netherlands, France and Belgium. Both Northern and Southern populations have a number of mosaic structures within *trmD*, providing evidence that recombination has occurred within this locus (Fig. 10, 13 and 15). Recombination within *gltA* cannot be established due to two one base pair substitutions separating the Northern and Southern populations. It is hypothesised that the division from the ancestral population occurred approximately 194 million years ago (MYA) by synonymous site analysis using *trmD*, although this estimate assumes no recombination. The high level of biogeographical structure presented in the spider mite population is very interesting and conflicts with the previous MLSA study in Chapter 3. It appears that biogeographical structure is present within *Wolbachia* spp., however this is only clear when examining closely related species and not on a Genus-wide scale. This is also telling in the context of the question asked in subsequent chapters, of whether *"Everything is everywhere"* (Beijerinck, 1913). Clearly this depends on phylogenetic scale, and the repeated isolation of clades from different hosts and different continents does not exclude the possibility of localised divergence on a much finer taxonomic scale.

iii) Host congruence v biogeographical structure: the true topology ?

The observed differences between loci in host adaptation and geographical structure are somewhat of a puzzle. Both phylogenies have multiple trees in agreement, and therefore are likely to reflect processes effecting the population rather than just a single gene. Co-evolution of spider mites and *Wolbachia* is implicated within *ftsZ* and *wsp*, but there is no evidence for this within the housekeeping genes where the isolates are divided according to geographical source. A possible explanation is outlined in Figure 19, where it is assumed that the housekeeping genes have been subject to local selective sweeps which have purged the phylogenetic signal evident at *wsp* and *ftsZ* A selective sweep is the reduction or elimination of variation among nucleotides as the result of recent and strong natural selection (Berry *et al.*, 1991; Page and Holmes, 2000). Strong selection for a single bacterial gene (a

selective sweep) may cause an entire chromosome to become predominant in the population, leading to a population wide reduction in genetic variation; a phenomenon known a genetic hitchhiking (Smith *et al.*, 2000). Alternatively, it may effect only the portions of the genome flanking the adaptive mutation (Guttman and Dykhuizen, 1994)

Genetic hitchhiking occurs when a neutral mutation "piggy-backs" with a mutation that natural selection has driven to fixation; i.e., as selection "sweeps" an advantageous mutation to fixation in a population, all those nuceotides linked to it – those parts of the genome that are not routinely separted from the selected site by recombination – are also swept to fixation (Nurminsky, 2001; Page and Holmes, 2000). The extent of reduced genetic variation within the sweep depends on the amount of subsequent recombination in the bacterial population (Saez *et al.*, 2003); a high recombination rate will keep the linkage block around the selected gene relatively small and lead to more chromosomal diversity away from the selected gene. Because of the hitchhiking of linked genes, it is generally difficult to identify which gene(s) has been the target of the selective sweeps. Normally, immediately after the sweep, little or no variation will be observed at either the target site or surrounding sites. As the distance increases from the target site the effect of the sweep decreases, resulting in higher diversity. *gltA* is the most conserved of the two housekeeping loci, therefore suggesting it is located near to a selective sweep on the genome, whereas *trmD* on the other hand has higher levels recombination, suggesting it is located further away from the selective sweep. The position of these loci on the *w*Mel genome places them on either side of *wsp*, which is conflicting with the evidence of a sweep as you would expect all the loci between and around *gltA* and *trmD* to have been affected. This may be explained by the extensive genome rearrangements *Wolbachia* spp. has undergone throughout evolution, and that the genes in the *w*Mel genome may not be in the same order as the *Wolbachia* sampled from the spider mites (Foster *et al.*, 2005; Wu, 2004). This presents a means of testing this hypothesis as it predicts that *gltA* and *trmD* should be linked in the genomes of the isolates infecting spider mites. If this prediction is fulfilled, this would represent the first example of a selective sweep on the basis of housekeeping gene sequences in *Wolbachia*, and one which is not apparent using the more commonly used *wsp* or *ftsZ* genes.

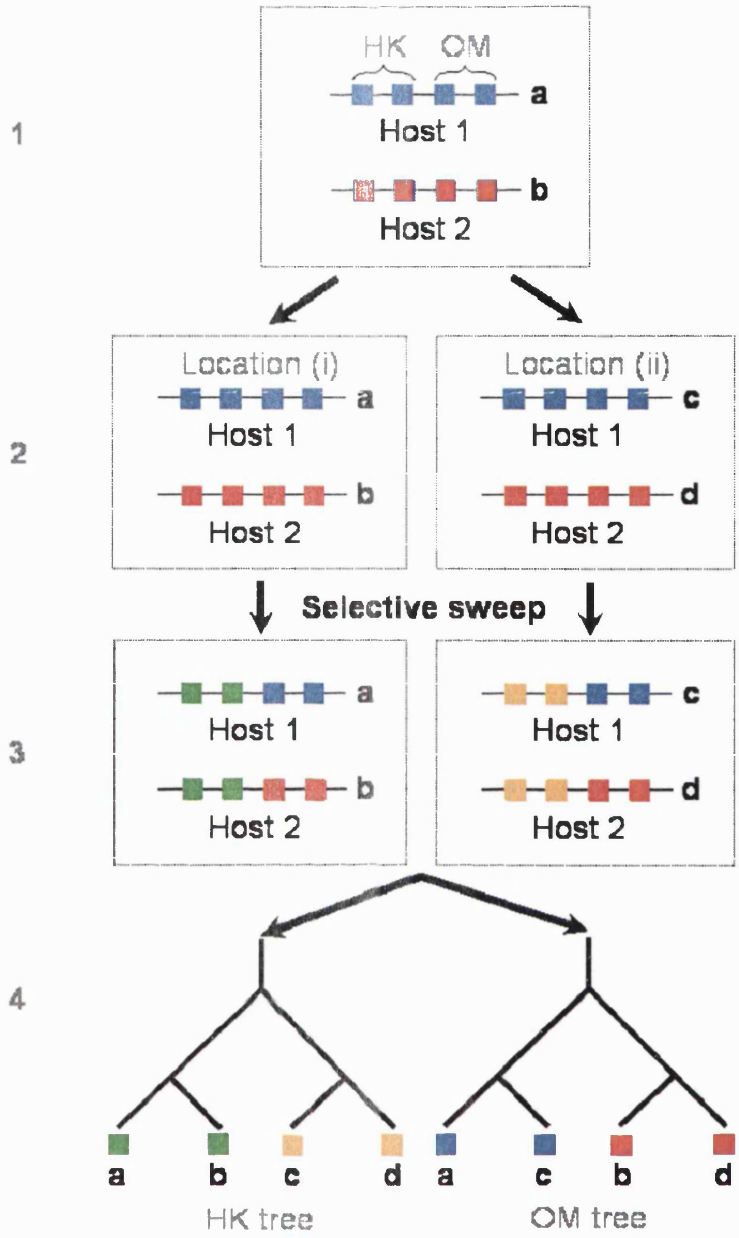



**Figure 19** Hypothetical evolutionary scenario to account for gene phylogeny corresponding to host adaptation in some genes and geographical source in others. Imagine the genome to consist of four genes, two of which are housekeeping (HK), and two associated with the outer membrane (OM). Initially these genes co-evolve (1) with different hosts to result in the red and blue genomes (a,b). This ancestral population is then split (2) into two distinct locales through migration or the emergence of a geographical boundary; each of the host adapted genomes is present in both locations (a-d). Within each of the geographical locations (3) the variation within the HK genes is purged independently by a selective sweep or genetic drift. This does not effect the variation within the OM genes, as the variation within these genes plays a role in host adaptation. Trees constructed (4) on HK genes will reflect geographical source, whereas trees constructed on OM genes will reflect the original divergence owing to host adaptation.

iv) Spider mite dataset is consistent with *Wolbachia* supergroup studies

Using *ftsZ*, the traditional choice of gene for the identification of *Wolbachia* supergroups, all twenty spider mite isolates were observed in *Wolbachia* supergroup B. The spider mite isolates were located in two main clusters in supergroup B, i) the first containing *B. kissophila* clades 2 and 3, in addition to the sexual spider mite isolate SM26, and ii) *B. kissophila* clade 1, the sexual spider mite isolate SM28 and the selected GenBank sequences.

v) Evidence for recombination with the spider mite dataset

As discussed, the general congruence between *wsp* and *ftsZ*, and the suggestion of host adaptation, is contradictory to the findings in Chapter 3 and challenges recent studies which have shown phylogenetic incongruence is the result of "pervasive" recombination (Baldo *et al.*, 2005a; Casiraghi *et al.*, 2005). Asexual reproduction of the host appears to limit the extent of recombination in *Wolbachia*. The two sexual spider mite species used in this study appear distinct at all loci, with the exception of *gltA*, and have undergone significant recombination events (e.g. Fig. 15), which is consistent with a sexually reproducing population. However, only one isolate of each was characterised, and a larger dataset is needed to robustly test the hypothesis that recombination is more frequent in *Wolbachia* isolated from sexual rather than asexual spider mite species.

As described in Chapter 3, a number of mechanisms exist for horizontal transfer to occur within the *Wolbachia* population, such as phages, transposable elements and multiple infections (Bordenstein and Wernegreen, 2004; Sanogo and Dobson, 2004; Wu, 2004). *Cardinium*, a relatively new reproductive parasite, has been observed in *Brevipalpus* spider mite species, where it has been observed to induce similar phenotypes to that of *Wolbachia*, to which it is unrelated (Groot and Breeuwer, 2006). The bacterium is located in the ovaries of its host, similar to *Wolbachia* (Zchori-Fein and Perlman, 2004). This bacterium may aid horizontal transmission between *Wolbachia* spp. within multiple infections and has been observed within several of the spider mite species included in this study (personal communication; Vera Ros).

Although mosaicism was observed throughout the spider mite population by visual inspection, many of the tests either detected little or no recombination within the population. Thus occasional recombinational replacements between distinct lineages are apparent, although the rate of recombination has not been high enough within the dataset as a whole for detection by tests such as Sawyer's Runs test. Furthermore, selective sweeps and subsequent genome rearrangement has led to many signals of recombination being masked and hidden by further recombination events. What is clear is that both vertical and horizontal transmission both play important a role in the evolution of *Wolbachia* spp.

# CHAPTER FIVE

## CHARACTERISATION OF THE NATURAL POPULATION OF *VIBRIO* SPP. FROM THE GOWER PENNINSULA

## 5.1 INTRODUCTION

### 5.1.1 *Vibrio* spp.

Members of the *Vibrio* genus are defined as Gram negative, asporogenous rods (Fig. 1a) which encompass a diverse group of heterotrophic marine bacteria (Brooks *et al.*, 1998). Many of the 64 *Vibrio* species currently recognised (Thompson *et al.*, 2004b) are highly abundant in aquatic environments, including estuaries, marine costal waters and sediments, and aquatic culture setting worldwide, due to their requirement for NaCl for optimal growth (Barbieri *et al.*, 1999; Heidelberg *et al.*, 2002; Singleton *et al.*, 1982). They are known to be commensal or pathogenic to corals (Banin *et al.*, 2000; Kuberski *et al.*, 1979), molluscs (Sawabe *et al.*, 2003), mussels and oysters (Kampelmacher *et al.*, 1972), crabs (Krantz *et al.*, 1969), seagrass sponges and shrimps (Johnson and Shunk, 1936; Verdonck *et al.*, 1997) and fish (Kothary and Kreger, 1985; Myhr *et al.*, 1991; Tison *et al.*, 1982) and can even form complex symbiotic associations with seaweed and phytoplankton (Nair and Simidu, 1987; Vugia *et al.*, 1997). The symbiotic colonisation of the squid *Euprymna scolope* by the bioluminescent species *V. fisheri* (Fig. 1b) is required for the normal development of the squid light organ (McCann *et al.*, 2003).

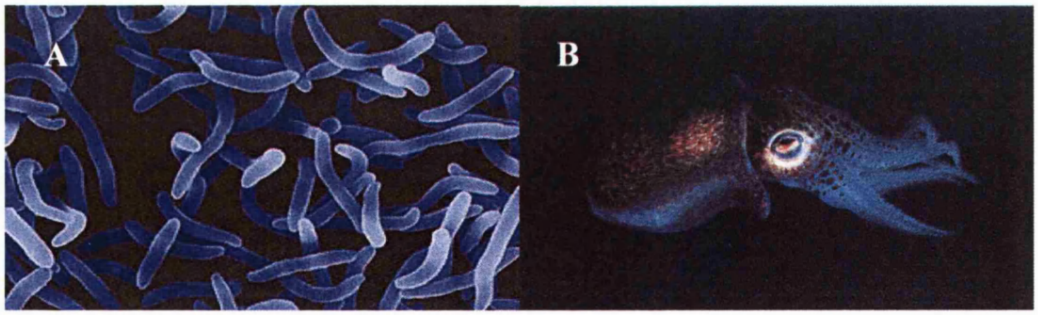



**Figure 1** A) *Vibrio cholerae* B) *Euprymna scolope,* the Hawaiian Bobtail squid, luminated by *V. fisheri.*

5.1.2 *Vibrio* spp. have two chromosomes

Five *Vibrio* genomes have been sequenced, representing four names species, *V. cholerae* 01 biovar El Tor N16961, *V. fisheri* ES114, *V. parahaemolyticus* RIMD 2210633, *V. vulnificus* CMCP6, *V. vulnificus* YJ016. All of these genomes consist of two circular chromosomes. Prokaryotic genomes often include extrachromosomal elements and *V. fisheri* ES114 and *V. vulnificus* YJ016 contain plasmids which are beneficial to the organism. Separate chromosomes, unlike plasmids are by definition indispensable because they harbour the essential genes necessary for survival. The genome of *V. cholerae* 01 biovar El Tor is 4,033,460 base pairs (bp) in length, of which chromosome I is 2,961,146 bp and chromosome II is 1,072,314 bp (Heidelberg *et al.*, 2000). Chromosome I contains most of the genes that are required for growth and pathogenicity, although some genes found on chromosome II are also required for normal cell function. However, chromosome II contains a large number of hypothetical genes and genes of unknown function (Heidelberg *et al.*, 2000).

The mechanisms underlying the unusual presence of two chromosomes in *Vibrio* are still unknown, however two hypotheses exist. The first and most likely, the plasmid hypothesis, suggests that chromosome II was originally a megaplasmid horizontally acquired by an ancestral *Vibrio* species (Egan *et al.*, 2005). As similarities are observed between putative essential genes on chromosome II of *V. cholerae, V, parahaemolyticus* and *V. vulnificus*, it is believed that speciation may have occurred after the initial acquisition of the chromosome II replicon, and acquisition of essential genes then occurred later and independently in the different species (Egan *et al.*, 2005). It is unclear why chromosome II has not integrated into chromosome I. One such reason may be that there is a selective advantage to having two chromosomes. For example, by reducing replicon size, two chromosome genomes are able to replicate at a much faster rate (the doubling time of *V. cholerae* in optimal medium is half that of *E. coli* K12), thereby providing the organism with a competitive advantage.

A second explanation is that chromosome I is exploited in response to environmental stimuli (Heidelberg *et al.*, 2000). Under extreme conditions, it may be favourable for one chromosome to partition to daughter cells in the absence of the other

chromosome, thus establishing itself in a "viable, but non-culturable" (VBNC) state, where the cells do not replicate or consume any nutrients. Although the cells are unable to grow in conventional media, they still maintain their viability, with key components of metabolism maintained while they are in this VBNC state (Gonzalez-Escalona *et al.*, 2006). VBNC *Vibrio* bacteria will be discussed later in this chapter.

### 5.1.3 Vibrios and aquaculture

As well as forming symbiotic relationships with various hosts, *Vibrio* bacteria have an important role in nutrient cycling in aquatic environments by utilising dissolved organic matter (Heidelberg *et al.*, 2000; Sherr and Sherr, 2002). As well as providing polyunsaturated fatty acids to other aquatic organisms, many of which are unable to provide these essential nutrients themselves, Vibrios are able to break down chitin, an amino sugar present in the ocean, and a homopolymer of *N*-acetyl-D-glucosamine (Cottrell *et al.*, 2000; Riemann and Azam, 2002). Additionally, *Vibrio* bacteria are producers of antibiotic compounds and selected species have been shown to reduce levels of other marine bacteria, such as α-proteobacteria and *Alteromonas* (Thompson *et al.*, 2004b; Long and Azam, 2001). The commensal relationships between Vibrios and chitinous zooplankton is critical for biodegeneration and biogeochemical recycling, and underlines their potential as valuable indicator species for monitoring pollution and the effects of climate change.

*Vibrio* infections are not limited to humans, various species of this genus are also devastating pathogens for animals reared in aquaculture, especially that of marine fish, shrimps and bivalves. The most common and well known infection is vibriosis, a systemic bacterial infection of primarily marine and estuarine fishes. A number of *Vibrio* species have been implicated as causative agents, including *V. anguillarum, V. salmonicida, V. ichthyoenteri* and *Photobacterium damselae* (Bullock, 1987). Vibriosis is characterised by a number of symptoms, including skin ulcers, septicaemia characterized by erythema, haemorrhaging, fin rot, anaemia and exophthalmia.

The overuse of antibiotics and especially quinolons is common in aquaculture, and has resulted in the development of multi-drug resistant strains of *Vibrio* bacteria, several of which have acquired resistance to the most commonly employed antibiotics leading to the generation of resistant strains of vibrios (Roque *et al.*, 2001). Alternatives to antibiotics as a control method are already in use. For example, vaccination of *Anguilla anguilla L.*, the European eel, against *V. vulnificus,* has already proved to be very successful in controlling infections (Esteve-Gassent *et al.*, 2003). Certain *Vibrio* strains exhibit probiotic effects and/or being symbionts of commercially important organisms (Thompson *et al.*, 2004b), recent studies have suggested that such strains could act as biocontrol agents, reducing the need for antibiotics in aquaculture (Verschuere *et al.*, 2000). Austin *et al.*, (1995) reported a probiotic strain of *V. alginolyticus* which did not cause any harmful effects in salmonids. Although the use of probiotics is encouraging in aquaculture, it is unknown if these bacteria are interacting with the remaining aquatic microbial flora. As seen with antibiotics, probiotic *Vibrio* spp. could have devastating effects on aquaculture if "virulent" recombination events were to occur, transforming the beneficial probiotic strain into a potentially lethal pathogen to aquaculture. These so-called probiotic bacterial strains could be fuelling the aquatic microbial food web (Thompson *et al.*, 2004b).

### 5.1.4 Vibrios and public health

Of the 64 *Vibrio* species documented, at least ten species are occasional food-borne pathogens and of medical importance to man, of which three species (*V. cholerae, V. parahaemolyticus, V. vulnificus*) are opportunistic pathogens associated with severe disease. *V. cholerae* is the causative agent of the severe life-threatening diarrheal disease cholera. To date, there have been 7 major pandemics since 1817, six of which were attributed to the classical biotype, with the most recent outbreak associated with the El Tor Biotype (Heidelberg *et al.*, 2000). *V. parahaemolyticus* and *V. vulnificus* together are responsible for most cases of seafood poisoning. *V. parahaemolyticus* is mostly prevalent in countries that have elevated levels of seafood consumption, such as Japan. Epidemiological and clinical investigations have shown *V. vulnificus* causes septicaemia, gastroenteritis and wound infections

originating from a marine environment. This organism is the major cause of death from eating raw oysters (Brooks *et al.*, 1998).

### 5.1.5 *Vibrio* ecology

Although much is known about *Vibrio* spp. when they cause disease, ecology of these species in the natural habitat is largely unknown. *Vibrio* populations are known to be sensitive to key environmental parameters as studies have shown striking seasonal fluctuation within aquatic environments with links to salinity and temperature (del Refugio Castaneda Chavez *et al.*, 2005; Jiang *et al.*, 2000b; Randa *et al.*, 2004). Salinity is an ecological factor of considerable importance, influencing the types of organisms that live in a body of water. It has been identified as playing a role in controlling *V. vulnificus* abundance, and although its precise effects are not clear, within *V. alguillarum* it is believed to affect the chemotactic response of the bacterium via the signal transduction pathway (Larsen *et al.*, 2004). Several studies have identified an inverse correlation between *V. vulnificus* abundance and salinity (Larsen *et al.*, 2004; Wright *et al.*, 1996), however others have suggested that there is an optimum salt concentration.

Similarly, temperature-mediated dynamics have been shown in several species of *Vibrio* in a positively-correlated manner (del Refugio Castaneda Chavez *et al.*, 2005; DePaola *et al.*, 2003; Randa *et al.*, 2004), thus total *Vibrio* concentrations are much higher during the warmer summer months than the cold winter months. *V. cholerae* and *V. parahaemolyticus*, two human pathogens which potentially occupy a similar ecological niche to that of *V. vulnificus*, also demonstrate seasonal variation in aquatic environments (DePaola *et al.*, 2003; Jiang and Fu, 2001; Larsen *et al.*, 2004), with up to a two-fold difference in total *Vibrio* concentrations observed between winter and summer (Thompson *et al.*, 2004c). Furthermore, studies have shown total *Vibrio* concentration is affected more by temperature than by salinity (DePaola *et al.*, 2003).

The effect of temperature on *Vibrio* populations has particular relevance to the seasonal occurrence of cholera outbreaks in many parts of the world. Historically,

cholera has been a problem in coastal cities, especially those where water quality is poor. It is part of the normal microflora in estuarine and marine environments (Colwell, 1996). The link between environmental *V. cholerae* and clinical cholera was made when the genetic relationship and diversity of *V. cholerae* isolated from both clinical and environmental sources indicated that some clinical strains were closely related to their environmental counterparts (Jiang *et al.*, 2000a). What is less well understood is what happens to the cholera population between epidemics, particularly in the winter months. Two possibilities seem likely; firstly, during these colder periods *Vibrio* bacteria enter a state known as Viable But Non-Culturable (VBNC) in which the bacteria become dormant, spore-like formations which are difficult to culture in the laboratory. At < 0.1 colony-forming units per millilitre (cfu $ml^{-1}$) this VBNC state can occur when temperatures drop below 13-15°C (Kaneko and Colwell, 1974). Secondly, host-associations may provide refugia in the winter months by attaching to the egg casing & gut of microscopic zooplankton, notably the copepods. These eggs are dispersed into the water (Dumontet *et al.*, 1996). Vibrios play a critical role in the breakdown of chitin, a component in the exoskeletal shell that is common to copepods (Colwell, 2000). *V. cholerae* is able to scavenge nutrients from the environment, due to duplication of genes coding for chitinase. Together with the phosphoenolpyruvate phosphotransferase system, *V. cholerae* produces and transports chitin-derived disaccharides released from the exoskeletons of zooplankton to which they attach (Schoolnik and Yildiz, 2000). During spring, when the water temperatures rise and phytoplankton blooms for the copepods to graze on, the bacteria become abundant, being "amplified", as the numbers of copepods increase. These studies show how an understanding of the natural ecology of cholera can allow the prediction of cholera outbreaks.

VBNC Vibrios can still cause disease even though the cells themselves cannot be cultured if the VBNC cells resume a vegetative state in the host. Culturable bacteria have been obtained following ingestion of VBNC cells 48 hours previously by human volunteers (Colwell, 2000). It is now hoped that by measuring water temperature, salinity and phytoplankton levels, scientists may be able to predict outbreaks and reduce the incidence of cholera in problem areas.

5.1.6 *Actinia equina* host population

A number of eukaryotic hosts will be sampled for this environmental study, however the primary host will be the sea anemone, *Actinia equina.* These primitive invertebrates belong to the Class Anthozoa (along with hard and soft corals), and are usually associated with a symbiotic Zooxanthallae (phytoplankton). It is impossible to determine age of a sea anemone, except for one that has been raised in an aquarium or tracked continuously in the wild from first settlement. Between 3cm and 6cm in diameter, *Actinia equina* (the Beadlet anemone) is very common along inter-tidal regions of the British Isles. Several features make these easily identifiable hosts ideal for this study. Firstly, they lack a shell or chitinous exoskeleton. The body is comprised of two layers, the ectoderm on the outside and the endoderm on the inside. Both of these layers are separated by the flexible jelly-like mesoglea, thus facilitating intrusive *in situ* swabbing with a cotton bud at low tide. In addition to being quick and easy, this does not require removal or damage of the host, and thus has minimal impact on the environment. Second, they have a very simple anatomy (Fig. 2), a single body cavity serves as a stomach, lung and intestine with a single opening serving both as a mouth, anus and presenting a convenient aperture for swabbing. Third, they are basically sedentary. Sea anemones often remain in same place for several days, weeks or even months, thus allowing re-sampling of a single host. They are not completely inactive and can move around on their pedal disc, so slowly that the movement can only be ascertained by a change of position in aquaria, or observed by time lapse photography. Fourth, they typically reproduce asexually, via pedal laceration, longitudinal fission, transverse fission, or budding. In pedal laceration, small sections of the pedal disk, where the anemone is attached, pinch off and grow into new anemones. In longitudinal fission, the anemone splits in two, with the plane of fission parallel to the animal's long axis running from its mouth to its foot. In transverse fission, the top section of the sea anemone pinches off. In budding, tentacles pinch off and develop into new animals. All these mechanisms of asexual reproduction result in local clonal populations. This then allows the potential for adaptive bacterial clones to widely colonise host of a given locale. However, they also retain the ability to occasionally reproduce sexually (Fig. 3), when the sperm from males enters the actinopharynx, where the egg is fertilized and then develops resulting in viviparous reproduction (Terrell, 2003). Therefore, geographically distinct populations are unlikely to be clone-

mates. Fifth, they have a broad diet, including zooplankton, small fish, shrimps and small shore crabs, meaning they will be exposed to a diverse reservoir of planktonic and host-associated Vibrios. Sixth, surprisingly, there are no reports in literature on the bacterial flora of sea anemones. This study will therefore provide novel data on this common inhabitant of the shore-line. Given the broad range of hosts with which *Vibrio* spp. are known to be associated, and the diet and ubiquity of sea anemones, they are almost certain to from intimate associations with these species.

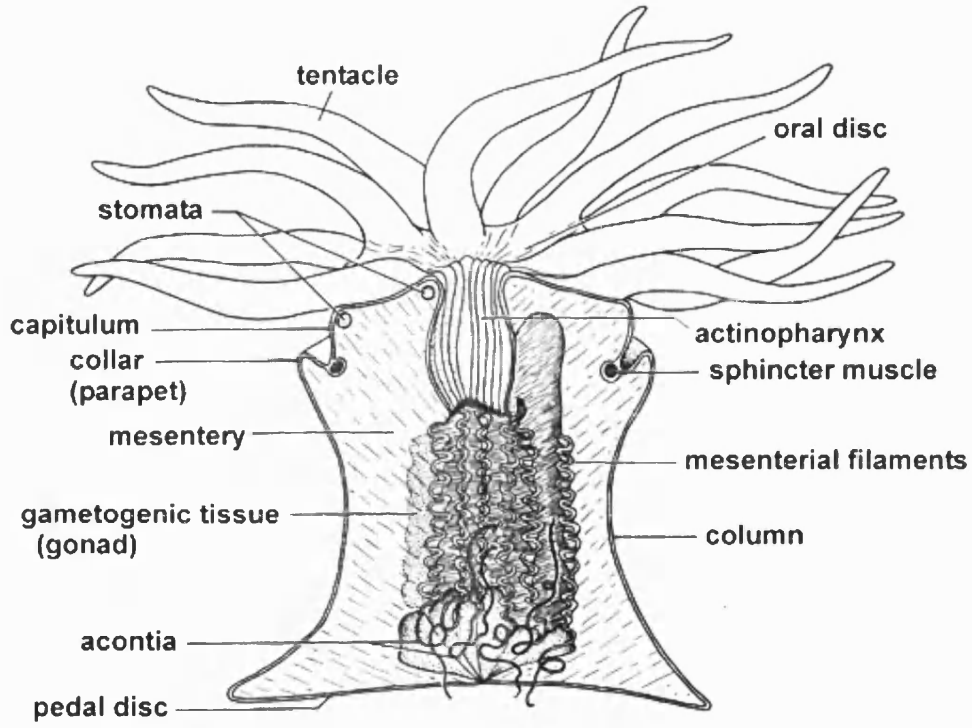



**Figure 2** Anatomy of *Actinia* sp.

**Figure 3** Young from a Beadlet anemone following sexual reproduction.

5.1.7 Sampling habitat

Sampling was conduced at the Gower, Wales, UK, a coastal peninsula surrounded by the Bristol Sea and Atlantic Ocean. With coastal cliffs, sand dunes, salt marshes, fresh water ponds/rivers, scrub and open commons on both acid and limestone soils, the Gower Peninsula is an ideal region to study the ecology and evolution of eukaryotic and prokaryotic species.

## 5.2 METHODS

### 5.2.1 Sampling

Sampling was conducted at Worms Head causeway, at the extreme South Western tip of the Gower Peninsula, Wales, extending to Rhossili Bay (Fig. 4). Two sampling sites were used (Fig. 5) approximately 1 km apart; the first at Rhossili bay on the South cliff face and the second at Worm Head Causeway (Fig. 6). Sampling was conducted a total of 4 times (Table 1).

| Sample site | Date | Location | Air temp | SST* |
|---|---|---|---|---|
| 1 | June 2004 | Worms Head Causeway | 20°C | 15°C |
| 2a | July 2004 | Rhossili Bay | 21°C | 21°C |
| 2b | July 2004 | Worms Head Causeway | 21°C | 21°C |
| 3 | April 2005 | Worms Head Causeway | 15°C | 13°C |
| 4a | August 2005 | Rhossili Bay | 20°C | 21°C |
| 4b | August 2005 | Worms Head Causeway | 20°C | 21°C |

**Table 1** Sample site information. *SST=sea surface temperature.

### 5.2.2 Isolation of Samples

Isolates from UK were recovered as described previously in Chapter 2, Section *2.2 Preparation and Storage of Cell and DNA stocks.* The target host organism of the UK sampling was *Actinia equina.* In addition to *A. equina,* isolates were also recovered from periwinkles (*Littorina* spp.), limpets (*Patella vulgata*), common shore crabs (*Carcinus maenas*), mussels (*Mytilus edulis*) and sea water. The number of host organisms sampled varied across each of the sample sites (Table 2).


**Figure 4** Location of the Gower Peninsula within Wales, UK.

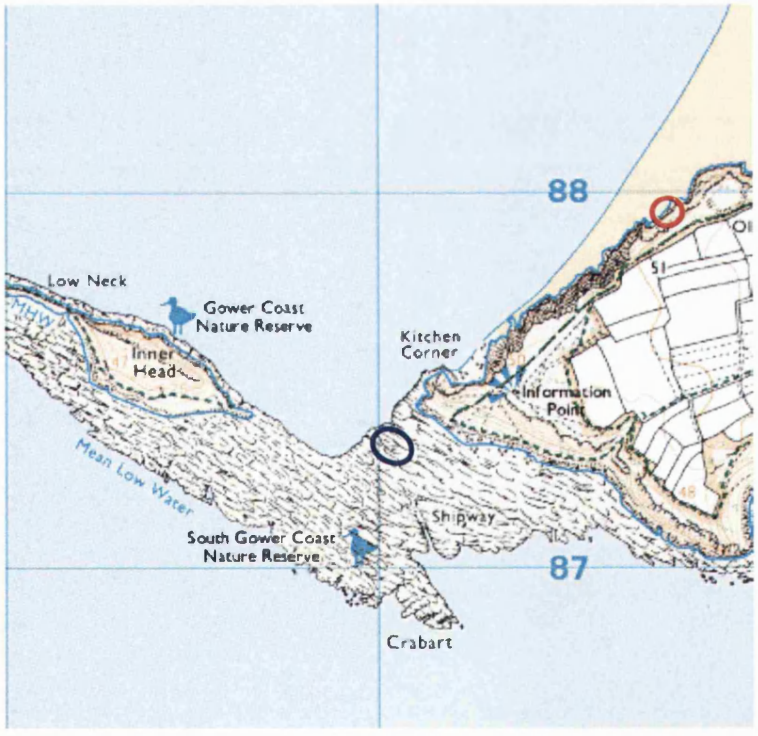



**Figure 5** Location of the two UK sampling sites. Rhossili bay is marked red and Worm Head Causeway is marked blue.

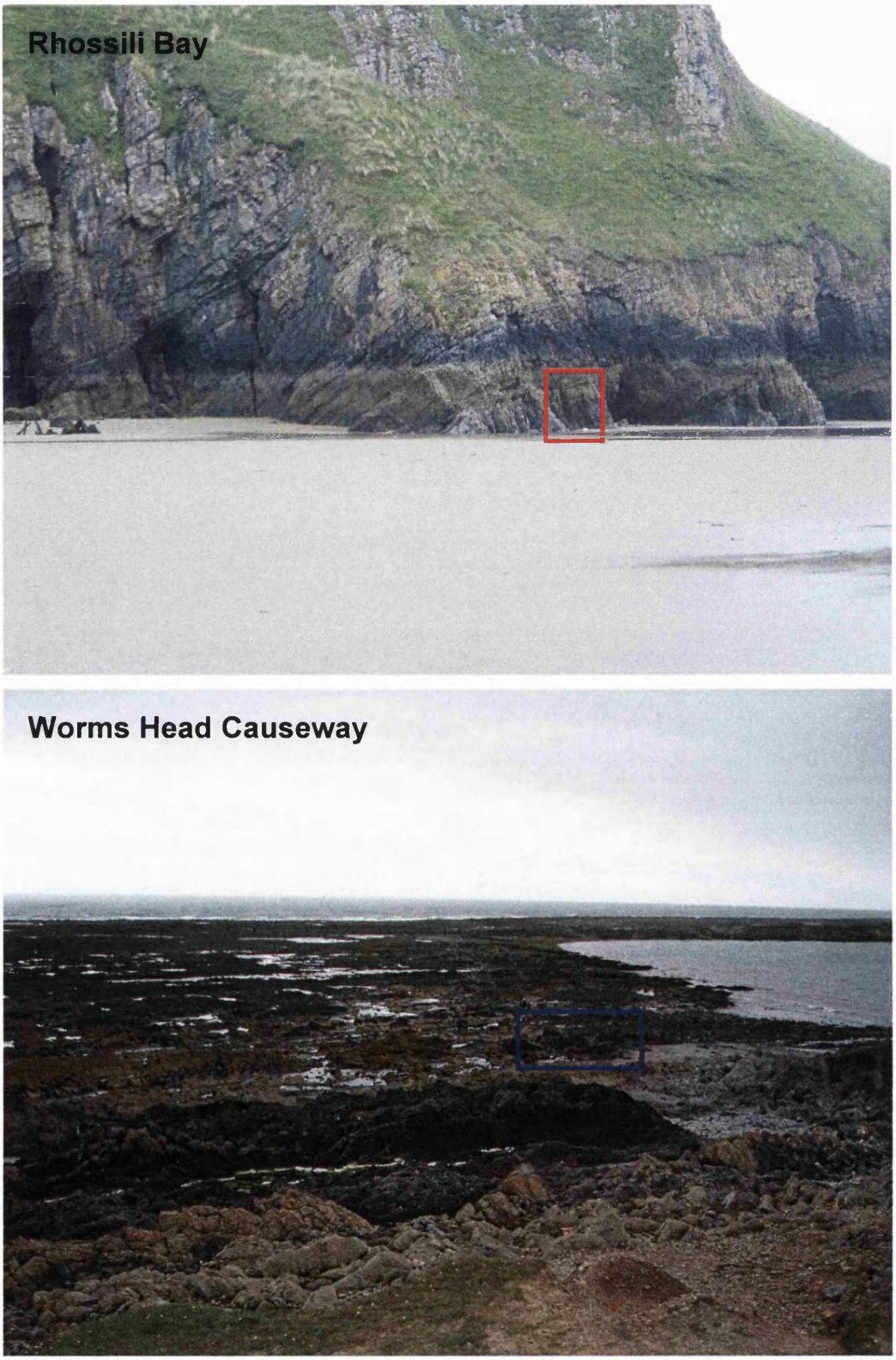



**Figure 6** Sampling locations. The area sampled at Rhossili Bay is identified by a red rectangle and Worm Head Causeway is marked by a blue rectangle.

| Sample date | Target organism | No. samples | % positive for *Vibrio* sp. |
|---|---|---|---|
| S1 June 2004 | Sea anemone | 2 | 100 % |
| | Periwinkle | 1 | 100 % |
| | Mussel | 2 | 100 % |
| | Crab | 1 | 100 % |
| | Limpet | 2 | 100 % |
| | Sea water | 1 | 100 % |
| S2 July 2004 | Sea anemone | 64 | 81.3 % |
| | Periwinkle | - | - |
| | Mussel | - | - |
| | Crab | - | - |
| | Limpet | - | - |
| | Sea water | 3 | 100 % |
| S3 April 2005 | Sea anemone | 10 | 90 % |
| | Periwinkle | - | - |
| | Mussel | - | - |
| | Crab | - | - |
| | Limpet | 4 | 100 % |
| | Sea water | - | - |
| S4 August 2005 | Sea anemone | 6 | 83.3 % |
| | Periwinkle | 19 | 94.7 % |
| | Mussel | - | - |
| | Crab | - | - |
| | Limpet | - | - |
| | Sea water | 13 | 100 % |

**Table 2** Percentage of eukaryotic hosts and sea water samples from which *Vibrio* spp. could be identified.

### 5.2.3 Isolate nomenclature

Each individual *Vibrio* isolate was given a unique four-part code which allowed its easy identification. For example, S2aSA80iv:

- S2a sample site (1, 2a, 2b, 3, 4a, 4b)
- SA target host organism
  - SA – Sea anemone
  - P – Periwinkle
  - M – Mussel
  - C – Crab
  - L – Limpet
  - S – Sea water
- 80 identifier for host organism
- iv identifier for strain

### 5.2.4 Choice of genes, primer design & PCR conditions

A total of 8 primer pairs were tested, of which 4 were used for the MLSA study in addition to 16S rRNA (Table 3). Within most MLSA schemes the loci are evenly spaced around the genome, however within this study the MLSA loci are clustered around one third of the genome (Fig. 7). It was decided gene function should have priority over gene location, in the hope of detecting large section genome replacements. PCR and sequencing was carried out as previously described (Chapter 2 Section 2.4 *Methods for Polymerase Chain Reaction and Sequencing*), except for *recG*, which incorporated a touchdown PCR procedure (Don *et al.*, 1991; Santos and Ochman, 2004) using the following conditions:

| | | | |
|---|---|---|---|
| Initial denaturation | | 2 mins | 94°C |
| 10 cycles: | denaturation | 1 min | 94°C |
| | annealing | 1 min | 60°C(-1°C/cycle) |
| | extension | 1 min | 72°C |
| 21 cycles: | denaturation | 1 min | 94°C |
| | annealing | 1 min | 50°C |
| | extension | 1 min | 72°C |
| Final extension | | 5 mins | 72°C |

PCR conditions for each loci varied by annealing temperature and extension time (Table 4). All primer sequences can be found in Appendix C.

| Loci | Gene function | Primer design method |
|---|---|---|
| *recA* | Classic "housekeeping" (MLST) gene. Involved in recombination & repair. | (Thompson *et al.*, 2004c) |
| *recG* | Involved in recombination & repair (Holliday junction migration). | (Santos and Ochman, 2004) |
| *mdh* | Catalyses the conversion of malate to oxaloacetate (essential metabolic gene) | Blastn |
| *ompK* | Codes receptor for broad host-range Vibriophage KVP40. | By eye |
| 16S rRNA | Involved in the process of translation and gene expression. | (Young *et al.*, 1991) |

**Table 3** Gene function & primer design method for each gene.

| Loci | Annealing | | Extension | |
|---|---|---|---|---|
| | Temperature | Time | Temperature | Time |
| *recA* | 50°C | 1 min | 72°C | 2 min |
| *recG* | touchdown | 1 min | 72°C | 1 min |
| *mdh* | 48°C | 1 min | 72°C | 1.5 min |
| *ompK* | 48°C | 1 min | 72°C | 1.5 min |
| 16S rRNA | 62°C | 1 min | 72°C | 1 min |

**Table 4** PCR conditions for each gene.

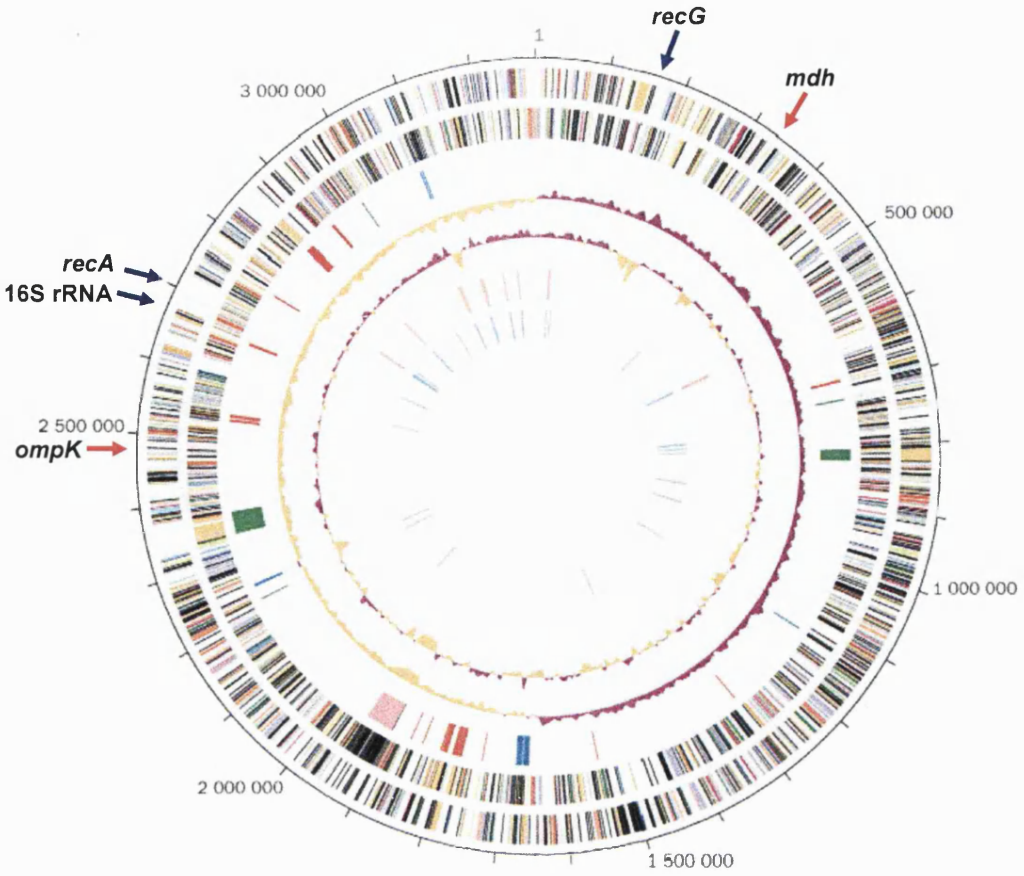



**Figure 7** *Vibrio parahaemolyticus* genome labelled with *Vibrio* MLSA genes. Genes located on the forward and reverse strand are coloured red and blue respectively. Modified figure (Makino *et al.*, 2003).

5.2.5 Nucleotide sequence analysis

-Sequence editing and alignment

As described in Chapter 2 Section 2.5 *Nucleotide sequence analysis*.

5.2.6 Phylogenetic analysis

-*recA* species assignment

Putative species assignments were assigned using the Basic Local Alignment Search Tool (BLAST) algorithm implemented at the NCBI website

(http://www.ncbi.nlm.nih.gov/BLAST). All *recA* sequences were compared to the Genbank database. The highest observed best bit score and E-values for each isolate were selected and the relevant species name(s) assigned on this basis.

-Distance methods

As described in Chapter 2 Section 2.6 *Phylogenetic Analysis.*

-Bayesian inference of phylogeny

As described in Chapter 2 Section 2.6 *Phylogenetic Analysis.*

The following parameters were used to create the Bayes block:

- number of generations = 6,000,000
- number of chains = 4
- sampling frequency = 1000
- print frequency = 100
- burnin = 20%.

### 5.2.7 Tests for Recombination

-Tests of neutrality

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Sawyer's Run Test

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Population-scaled recombination rate (ρ)

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-DnaSp

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Splits decomposition

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

5.2.8 Other methods of analysis

-eBURST

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

-Datamonkey

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

## 5.3 RESULTS

### 5.3.1 The dataset

A total of 330 strains were isolated and sequenced at 5 loci, however not all strains were sequenced at each loci. The total number of isolates varied with each gene; 16S rRNA (n=296), *recA* (n=281), *recG* (n=204), *mdh* (n=280), *ompK* (n=271). Overall population analysis was conducted with all nucleotide sequence information available. 164 strains were sequenced at all genes and in these cases the sequences were concatenated for further analysis (see Appendix C for further details). *ompK* was the most variable of the genes sequenced, and contained a large number of indels (Table 5).

| Locus | $\pi$ |
|---|---|
| 16S rRNA | 0.032 |
| *recA* | 0.047 |
| *recG* | 0.094 |
| *mdh* | 0.050 |
| *ompK* | 0.157 |

**Table 5** Average pairwise percentage sequence difference for each loci.

### 5.3.2 Species assignments

Previously, the majority of phylogenetic analysis of *Vibrio* spp. has been performed with *recA*, a typical housekeeping gene, ubiquitous to bacteria and subject to strong purifying selection due to its vital roles in DNA recombination and repair. We therefore used this gene as a "benchmark" for assigning species using the data of Thompson *et al.* deposited on GenBank (Thompson *et al.*, 2004a). We were therefore able to explore the extent to which *recA* sequence alone could reliably identify species. Species assignments were made on the basis of Blastn similarity (E) scores (Altschul *et al.*, 1990) for each query *recA* sequence against all the *Vibrio* *recA* sequences in GenBank (data not shown). A total of twelve *Vibrio* species were identified using this approach (Table 6).

| Species identified | % of species isolated |
|---|---|
| *V. cyclitrophicus* | 34.5 |
| *V. splendidus* | 26.3 |
| *V. kanaloae* | 12.1 |
| *V. diabolicus* | 9.6 |
| *V. tasmaniensis* | 9.3 |
| *V. pomeroyi* | 2.5 |
| *V. harveyi* | 2.5 |
| *V. ichthyoenteri* | 1.4 |
| *V. lentus* | 0.7 |
| *V. pacinni* | 0.4 |
| *V. tubiashii* | 0.4 |
| *P. eurosenbergii* | 0.4 |

**Table 6** *Vibrio* species identified according to *recA* species assignments. Based on 281 *recA* sequences available.

5.3.3 Sequence analysis

The alleles defined for the study were between 251 bp (16S rRNA) and 660 bp (*recA*) in length, and between 43 (*mdh* & 16S rRNA) and 58 (*recG*) alleles were present per locus (Table 7).

| Locus | Fragment size (bp) | No. of alleles | No. of variable sites | % Variable sites | dS/dN |
|---|---|---|---|---|---|
| 16S rRNA | 251 | 43 | 72 | 28.7 | - |
| *recA* | 660 | 45 | 184 | 27.9 | 45.50 |
| *recG* | 348 | 58 | 181 | 52.0 | 24.22 |
| *mdh* | 468 | 43 | 208 | 44.4 | 24.25 |
| *ompK* | 576 | 55 | 335 | 58.2 | 3.73 |

**Table 7** Genetic diversity of *Vibrio* spp. MLSA loci (n=164).

In order to gauge the selective pressure of the protein-coding genes, the dS/dN ratio was calculated as described on within Chapter 2. *recA* is the most evolutionary conserved protein-coding gene in that it exhibits both the highest dS/dN ratio (45.50; Table 7) and the lowest average pairwise divergence ($\pi$; Table 6). It is therefore under the greatest level of purifying selection of the four genes. *mdh* and *recG* exhibit high dS/dN ratios (24.25 and 24.22 respectively), almost half of that of *recA,* however *mdh* has a similar average pairwise divergence ($\pi$) to that of *recA*. *recG*

instead has a higher average pairwise divergence (9.4%) in addition to the highest percentage of polymorphic sites and the most alleles for the three housekeeping genes. Therefore, *recG* is the most divergent of the three genes, although it is still under purifying selection. The lowest dS/dN ratio is observed with *ompK* (3.73), indicating it has a higher proportion of non-synonymous changes and the lowest level of purifying selection of the four loci. *ompK* is the only gene to contain large indels (three in total), one of which spans 15% of the maximum allele length. The presence of indels, the highest average pairwise divergence (15.7%) and the greatest percentage of polymorphic sites, all support the observation that *ompK* is the most variable and diverse of the four loci. Although no positively selected sites could be identified within the gene (analysis by Datamonkey, not shown), it is clear that *ompK* is under lower purifying selection within the gene as a whole. This is consistent with its role as a phage receptor, in which it is expected to be under rapid evolution due to its relationship with phage on the environment.

### 5.3.4 Phylogenetic analysis

Phylogenetic trees were constructed independently for each locus (Fig. 8) using the MEGA suite of programs (Kumar *et al.*, 2004), version 3. The Neighbour-Joining method with the nucleotide Kimura 2-parameter correction was used for all strains, as more sophisticated approaches such as MrBayes were too computationally intensive for this large dataset. All trees were unrooted.

Neighbour-Joining trees 16S rRNA, *recA*, *recG*, *mdh* and *ompK* from all isolates where sequences were available are shown in Figure 8, sections a, b, c, d & e respectively. The trees are colour coded according to species assignments based on *recA* Blastn scores. Comparison of the phylogenetic trees provided further evidence differences in diversity between the gene loci.

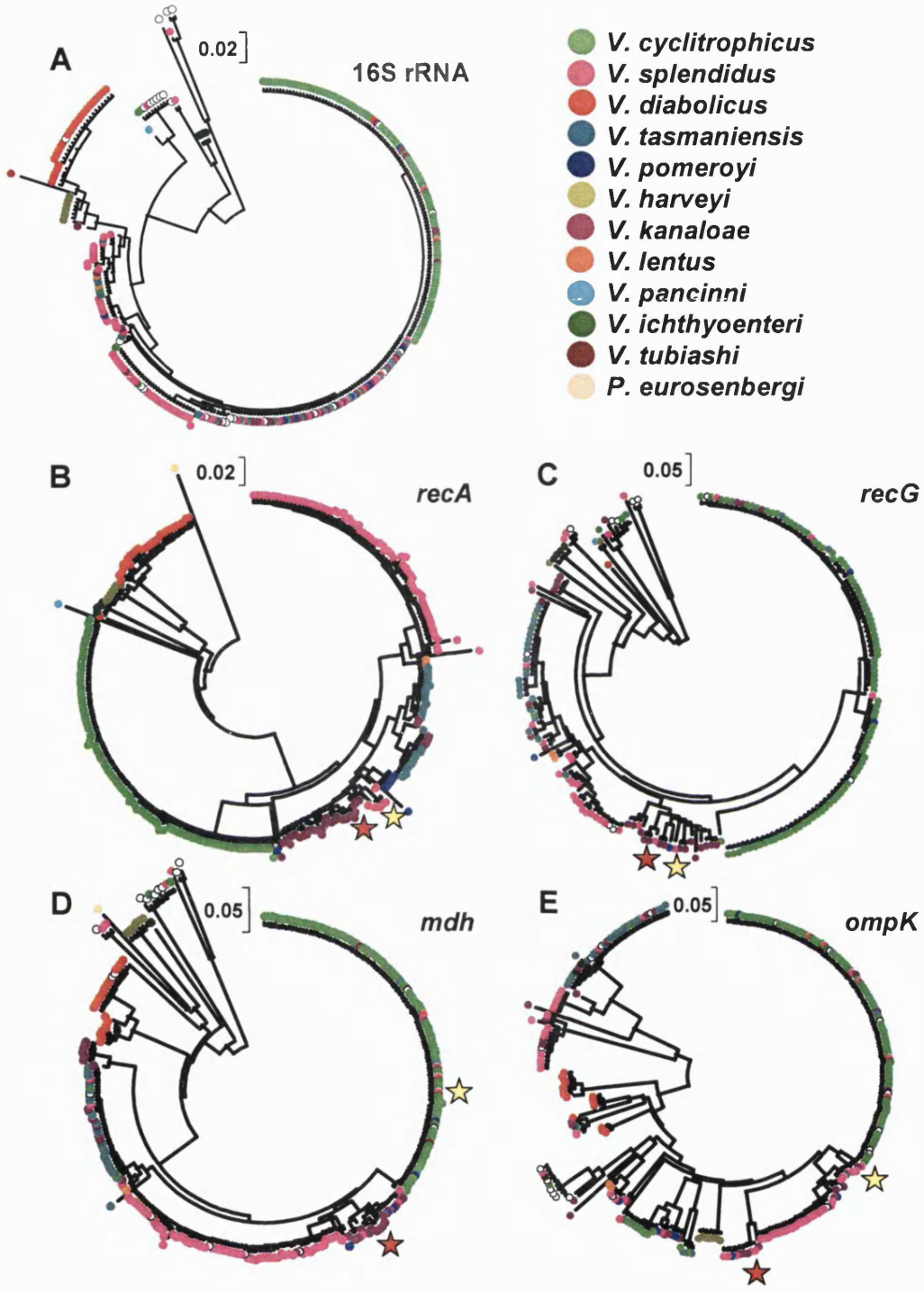



**Figure 8** Neighbour-joining trees for nucleotide sequence of 16S rRNA and MLSA loci using all available nucleotide sequences. Each isolate is colour coded according to its *recA* species assignment. Open circles represent isolates which have no *recA* sequence data and therefore cannot be assigned to a *Vibrio* species. Red and gold stars are explained within the text.

Within *recA* (Fig. 8b), the majority of the tree is dominated by *V. cyclitrophicus*, which forms a single large cluster (34.5 % of the isolates, Table 6). *V. splendidus* forms two clusters, the largest of which (n=68) forms the second biggest species cluster on the tree, and a smaller cluster of 6 isolates closely associated with *V. kanaloae* and *V. pomeroyi*. Two *V. lentus* isolates group with the main V. splendidus cluster. At the bottom of the *recA* tree, three *Vibrio* species each form distinct, single species clusters (*V. diabolicus, V. harveyi* & *V. ichthyoenteri*), in addition to three, diverse *Vibrio* species (*V. pacinni, V. tubiashii* and *P. eurosenbergii*), each present as only a single copy and occupying separate branches to the rest of the isolates. *V. pomeroyi* forms a single cluster within the centre of the tree, with one isolate (S2bA29vi) showing more diversity within the group. Additionally, a single *V. pomeroyi* isolate (S2bA48v) does not group with the main cluster, but instead shows some similarity to the *V. cyclitrophicus* cluster. *V. tasmaniensis* and *V. kanaloae* are each separated into two separate clusters occupying similar areas of the tree. *V. tasmaniensis* is divided in two by the smaller of the two *V. kanaloae* clusters.

Both *mdh* and *recG* tree (Fig. 8c & 8d) has the strongest phylogenetic concordance in relation to *recA*. *V. cyclitrophicus* and *V. splendidus* both dominate the *mdh* tree forming 2 large clusters, in which there are a number of *Vibrio* isolates assigned to other *recA* designated species groups. *V. diabolicus* and *V. harveyi* again correspond to two discrete clusters, although appear less closely related than at *recA*. The three *V. pomeroyi* strains do not cluster, suggesting that this does not correspond to a discrete species, or that these strains have been misassigned. All the *V. tasmaniensis* cluster together, but five out of the six *V. kanaloae* isolates also cluster with this species as observed on the *recA* tree. *V. kanaloae* is divided into two clusters, however these are not as closely related as with *recA*.

From the remaining phylogenetic trees, 16S rRNA (Fig. 8a) shows the poorest phylogenetic resolution, thus making it relatively uninformative compared to the other loci. *V. harveyi, V. ichthyoenteri, V. diabolicus, V. cyclitrophicus* and *V. splendidus* each form separate species clusters on the tree, however, there is some evidence of recombination within the latter two clusters, as a number of *Vibrio* species are present within these groups. As discussed in the next chapter, the

clustering of strains from other species within the *V. cyclitrophicus* clade is most likely to result from recombination at *recA*, leading to species misassignments. For example, *V. diabolicus* (n=1), *V. pomeroyi* (n=1), *V. splendidus* (n=2) and *V. kanaloae* (n=6) are observed within the *V. cyclitrophicus* cluster. A large cluster in the centre of the tree contains a number of different *Vibrio* species; *V. splendidus* (n=13), *V. pomeroyi* (n=3), *V. kanaloae* (n=20), *V. tasmaniensis* (n=15), *V, cyclitrophicus* (n=1), *V, diabolicus* (n=1). Dominated by *V. kanaloae* and *V. tasmaniensis*, this cluster illustrates the poor discriminatory power of 16S rRNA for species identification.

*recG* appears to be a moderately divergent gene, despite the fact that it is a universally conserved status (Santos and Ochman, 2004). Firstly, the *recG* tree has the smallest data set of all the loci (n=204) and is underrepresented by some *Vibrio* strains, namely *V. diabolicus, V. splendidus, V. harveyi* and *V. pomeroyi. V. tasmaniensis* forms a single cluster of strains with relatively few isolates (n=3) clustering with other species around the tree. *V. splendidus* also forms a single cluster on the *recG* tree, as with *mdh*, in which few other species are found. *V. kanaloae* is again divided into two clusters, however these are not a closely related as with *recA* due to their greater distance apart on the tree. Interestingly, *recG* divides *V. cyclitrophicus* into two closely related clusters. An analysis of the *recG* genes from single representative strains from each of these clusters reveals that these sequences differ by twenty nucleotide changes (Fig. 9). These differences are non-randomly clustered; atypical *V. cyclitrophicus* isolates are identical to *V. splendidus* isolates for the first 102 bases which is strong evidence for mosaicism resulting from intragenus (inter-species) recombination.

*ompK* is by far the most divergent of the loci. Clustering according to *recA* species designations is observed with a number of the *Vibrio* species and these clusters are broken down into smaller groups, often located on different areas of the tree. For example, the *V. cyclitrophicus* cluster is observed as two groups. These groups correspond to different strains than the two *V. cyclitrophicus* groups noted for *recG* and are separated by *V. splendidus* and *V. harveyi* clusters. The *V. cyclitrophicus* smaller cluster (n=9) is located in a poorly resolved area of the tree, in which there appears to be no obvious species clustering. *V. tasmaniensis* forms a single cluster

of isolates at the most basal branch of the tree. A single grouping of *V. harveyi* is observed in the centre of the tree, and is no longer closely associated with *V. diabolicus*, in contrast to all other loci. *V. splendidus* is divided into two unrelated clusters, each located on different sections of the tree. The *Vibrio* species showing the least species cohesion on the *ompK* tree is *V. kanaloae*. As observed with previous loci, *V. kanaloae* formed two clusters, one of which was closely associated with *V. tasmaniensis*. However, on the *ompK* tree, very little *V. kanaloae* clustering is observed with the largest grouping totalling three isolates.

```
                                    1111 1111111112 222222222
                      14455666 7777890334 6778888990 223456778
                     3922814369 0258432251 8470123584 891054065
#S1S18iv             CGTATGACCG TGTGACTTGA CCGCGTTTTA TTGAACATG
#S4aA80iv            .......... .......... .......... .........
#S4bP18iii           .......... .......... .......... .........
#S1C01iv             GACGGAGATA CACAGTC... .......... .....TGC.
#S2bA52viii          GACGGAGATA CACAGTC... .......... .....TGC.
#S2aA29viii          GACGGAGATA CACAGTCCTG TTAAACACCG GCAGCTGCA
#S2aA29x             GACGGAGATA CACAGTCCTG TTAAACACCG GCAGCTGCA
```

**Figure 9** Evidence of mosaicism within *recG* resulting from an intragenic recombination event.

Although all the genes show a level of consistency concerning species assignments with the *recA* tree, several interesting and dramatic exceptions are noted. For example, strain S2aA03ii (highlighted in Figure 8 by a red star) has been assigned as *V. kanaloae* on the basis of its *recA* sequence, and for the loci *mdh* and *recG*, this isolate is observed in the *V. kanaloae* clusters on these trees. However, S2aA03ii has acquired a divergent *ompK* allele showing identity and clustering with *V. splendidus*. A second example of recombination between species is with S2bA51iii (highlighted in Figure 8 by a yellow star), assigned as *V. splendidus* on the basis of the *recA* sequence. This strain is identical to strains assigned as *V. kanaloae* at both *mdh* and *recG*, however, at *mdh* and *ompK* has acquired a divergent *V. cyclitrophicus* allele.

In addition, the genes vary in their ability to cluster named species. Although 16S rRNA performs well in clustering most species, there remains a large number of isolates belonging to multiple named species by *recA* which are clustered together at 16S rRNA. Conversely, the *ompK* tree tends to divide named species into distinct clusters in opposite sides of the tree. The three housekeeping genes, *recA, mdh* and *recG,* fall between these extremes. The cohesion of the named species varies according to species, as well as to gene loci. *V. kanaloae* in particular only forms a single cluster at *recA* (the basis upon which this species was identified), however none of the other loci strongly support this species assignment.

### 5.3.5 Concatenation of MLSA loci resolves phylogenetic inference

As described in Chapter 1, the concatenation of nucleotide sequences can result in a more robust phylogeny combining the signal from all loci. Nucleotide sequences of the four gene fragments were concatenated in-frame using a PERL script (kindly provided by Kevin Balbi) and a Neighbour-Joining tree was constructed (Fig. 10a) Two clearly defined species groupings can be seen at opposite ends of the tree (*V. cyclitrophicus* & *V. tasmaniensis*), however the central section of the tree is less clearly resolved and does not conform to previously observed *Vibrio* species groupings. Highly variable genes will contribute a disproportionate signal to the tree, and this poorly linked switching reflects bias towards the *ompK* gene. If *ompK* is removed from the concatenated data (Fig. 10b), five well-resolved species groupings are observed, although as discussed previously, *V. kanaloae* is the least cohesive. Concatenation therefore remains a useful approach, with the caveat that all genes make roughly equal contributions to the tree.

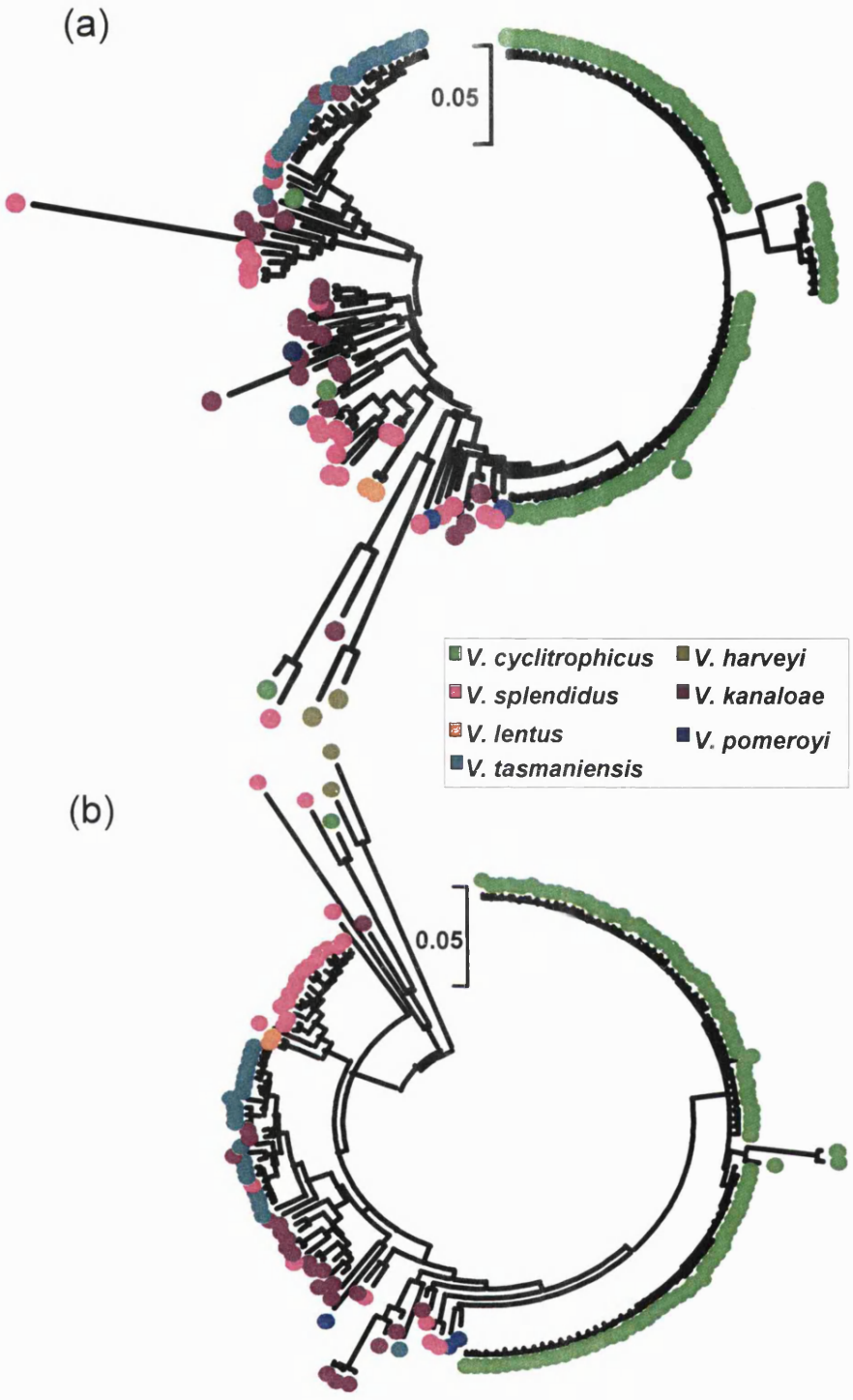



**Figure 10** MLSA loci provide greater phylogenetic resolution when nucleotide sequences are concatenated (n=164). a) Represents 4 MLSA loci, b) represents *recA, recG* & *mdh* MLSA loci.

5.3.6 Model of phylogenetic analysis

To assess the reliability of the Neighbour-Joining tees constructed, thirty strains were randomly selected (Appendix C) from the four loci concatenated MLSA tree (Fig. 10a). Bayesian analysis was selected as the most suitable method for the estimation of reliability of phylogeny. Bayesian inference of phylogeny is based upon the posterior probability distribution of trees, using Baye's theorem. Concatenated Neighbour-Joining trees were constructed using MEGA v3.1 and Bayesian trees by MrBayes v3.1.2, as previously described in Chapter 2 section 2.6 *Phylogenetic analysis.* To allow comparison, the Bayesian and Neighbour-joining trees were colour coded according to *recA* species assignments. Clustering according to *recA* species assignment is observed on both Neighbour-Joining and Bayesian trees. Again this is more evident when *ompK* is removed (Fig. 11b i & ii). Although the *Vibrio* species form clusters at different locations on the tree, in general the same isolates group together on both Neighbour-Joining and Bayesian trees. For example, within the *V. tasmaniensis* group, five *Vibrio* isolates (S4bP14iv, S2bA48iii, S2bA53vi, S2aA05ix, and S2aA30viii) repeatedly cluster together on all the trees. However the relationships between the clusters are poorly supported and differ according to phylogenetic method. Therefore, although this more detailed phylogenetic analysis provides strong support for clusters generally corresponding to named species, the concatenated data does not resolve the relationships between the species. This is likely to reflect the fact that the ancestral genome of each species was a mosaic from many different lineages. When *ompK* is included the bootstrap (or posterior probability) scores are lower than when this gene is removed again indicating the conflicts in the phylogenetic signal. Interestingly, in terms of grouping species assigned on the basis of *recA*, the Bayesian approach does not perform noticeably better than the far simpler Neighbour-Joining approach. For example, in Figure 11bi, the *V. cyclitrophicus* isolates, with the exception of S1L02v, cluster together to form one group, however, in Figure 11bii, this larger *V. cyclitrophicus* group is broken down into four groups of closely related isolates.

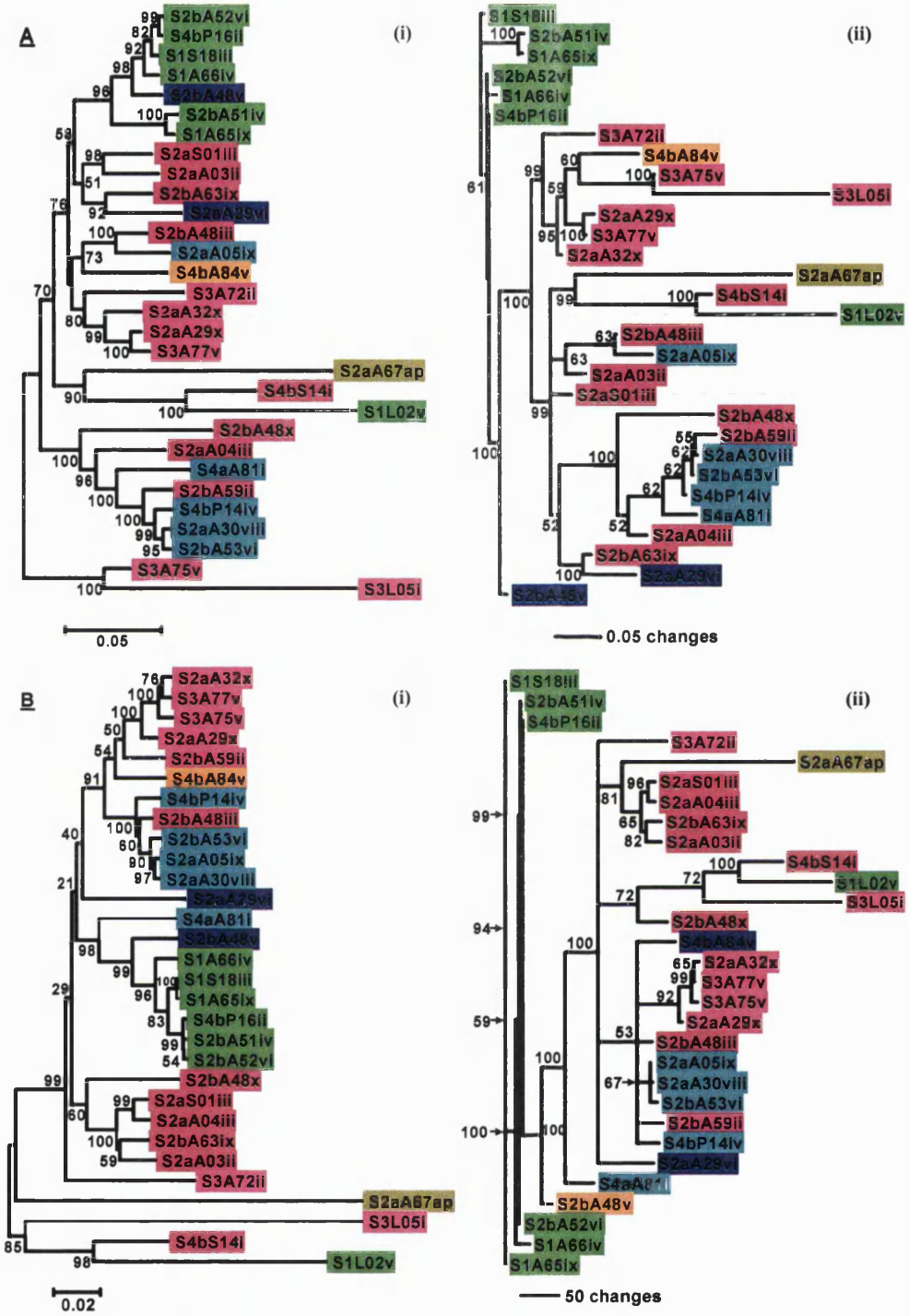



**Figure 11** Phylogenetic trees based upon the Neighbour-Joining & Bayesian method, using concatenated nucleotide sequences (n=30). Isolates are colour coded according to *recA* species assignments. A) Refers to 4 loci MLSA concatenated sequences, B) to 3 loci MLSA concatenated sequences, i.e. *recG, mdh* & *recA*. i) Refers to Neighbour-Joining method, ii) to Bayesian method with trees constructed in Paup*.

5.3.7 Evidence of host adaptation

A number of eukaryotic hosts were samples in the study to ascertain the relationship between host and *Vibrio* spp. isolated. Using the four loci MLSA concatenated Neighbour-Joining tree, colours were assigned to each eukaryotic host (Fig.12). *A. equina* clearly dominates the concatenated tree (77 %). There is no clear evidence of host adaptation corresponding to genotype, for example isolates recovered from *Actinia, Littorina, Patella* were all represented in the large *V. cyclitrophicus* cluster. By examining all isolates for which *recA* species assignments are available (n=281), rather than strains for which all three genes are available, a more detailed understanding of host adaptation can be observed (Table 8). The greatest host diversity is observed within *V. cyclitrophicus*, being isolated from all six "hosts", with the majority of this species was found in sea anemones (n=64), with significantly lower numbers (< 10) observed in the remaining hosts. The greatest species diversity is within sea anemones, with all of the twelve *Vibrio* species isolated recoverable from sea anemones. This difference in host and species diversity within sea anemones and *V. cyclitrophicus* is highly significant ($P < 0.0001$), indicating a departure from homogeneity between sea anemones and the remaining hosts, and between *V. cyclitrophicus* and the remaining eleven *Vibrio* species (Table 9). *V. diabolicus* was isolated from all but one of the hosts (limpets), with the majority of isolates observed in periwinkles and sea water. Of these, only one isolate was recovered from both sea anemones and mussels. The single *V. diabolicus* isolate from sea anemone may have resulted from environmental contamination, thus this result implies that *V. cyclitrophicus* is more commonly associated with sea anemones than *V. diabolicus*, given their relative frequencies in the sea water. However, the high level of *Vibrio* diversity observed within *Actinia* (see below) suggest that this species maybe frequently colonised by a large number of species.

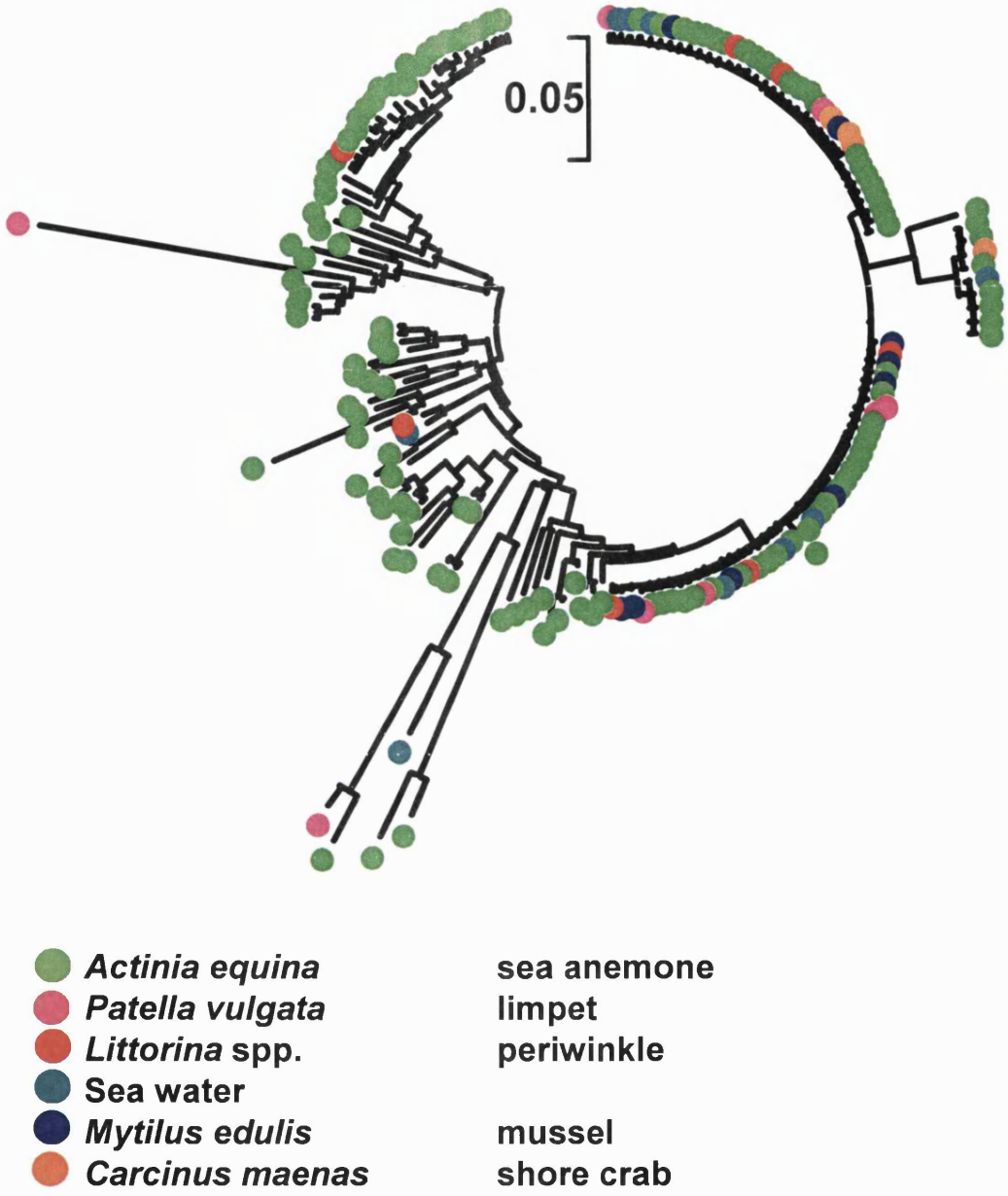



**Figure 12** Phylogenetic tree based on the Neighbour-Joining method, using 4 loci MLSA concatenated nucleotide sequences (n=164). Isolates are colour coded according to the eukaryotic host they were isolated from.

*V. splendidus* was recovered from four eukaryotic hosts, with 95 % of all these strains isolated from sea anemones. The remaining 5 % of this species was isolated from limpets (n=2), shore crabs (n=1) and periwinkles (n=1). *V. kanaloae* was predominantly isolated from sea anemones (n=28), however, was also recovered from periwinkles (n=2) and sea water samples (n=4). *V. tasmaniensis* and *V. pomeroyi* were recovered from only two eukaryotic hosts, both of which were common to sea anemones (n=25 & n=6, respectively). *V. tasmaniensis* was also isolated from periwinkles (n=1) and *V. pomeroyi* from a shore crab (n=1). The lowest level of host diversity was observed with *V. lentus* (n=2), *V. harvey*i (n=7), *V. ichthyoenteri* (n=4), *V. pacinni* (n=1), *V. tubiashii* (n=1) and *P. eurosenbergii* (n=1); these species were all recovered only from sea anemones.

| *Vibrio* species | *Actinia equina* | *Patella vulgata* | *Carcinus maenas* | *Mytilus edulis* | *Littorina* spp. | Sea water |
|---|---|---|---|---|---|---|
| *V. cyclitrophicus* | 64 | 7 | 5 | 9 | 5 | 7 |
| *V. tasmaniensis* | 25 | - | - | - | 1 | - |
| *V. kanaloae* | 28 | - | - | - | 2 | 4 |
| *V. splendidus* | 70 | 2 | 1 | - | 1 | - |
| *V. lentus* | 2 | - | - | - | - | - |
| *V. pomeroyi* | 6 | - | 1 | - | - | - |
| *V. harveyi* | 7 | - | - | - | - | - |
| *V. diabolicus* | 1 | - | 2 | 1 | 10 | 13 |
| *V. ichthyoenteri* | 4 | - | - | - | - | - |
| *V. pacinni* | 1 | - | - | - | - | - |
| *V. tubiashii* | 1 | - | - | - | - | - |
| *P. eurosenbergii* | 1 | - | - | - | - | - |

**Table 8** Number of *Vibrio* spp. recovered from each eukaryotic host and sea water (n=281). *Actinia equina* = sea anemone; *Patella vulgata* = limpet; *Carcinus maenas* = shore crab; *Mytilus edulis* = mussel; *Littorina* spp. = periwinkle.

| | | Eukaryotic host | |
|---|---|---|---|
| | | *A. equina* | Other |
| Species | *V. cyclitrophicus* | 58 | 33 |
| | Other *Vibrio* spp. | 59 | 5 |

**Table 9** Chi-squared test ($P < 0.0001$) indicates a departure from homogeneity between the 2 categories; 1) between eukaryotic hosts and 2) between *Vibrio* species.

5.3.8 No variation is observed between sampling sites

To address the extent of local adaptation within the two sampling sites, the four loci concatenated Neighbour-Joining tree was colour coded (Fig. 13) according to Rhossili Bay and Worms Head causeway. No differences could be identified between the sampling sites, indicating that the two sites are most likely colonised by the same *Vibrio* species; they should in fact be considered one sampling site. To measure the extent of population subdivision between the two sites, Wrights inbreeding coefficient ($F_{ST}$) was calculated for each of the coding loci (Table 10). High levels of gene flow are observed between the two sites, more so with *recA* and *recG*, than *mdh* and *ompK*. Therefore, there is little geographic structure between Rhossili Bay and Worms Head causeway, again suggesting that the sites are similarly colonised.

| Locus | $F_{ST}$ value |
|---|---|
| *recA* | 0.01415 |
| *recG* | 0.01140 |
| *mdh* | 0.02291 |
| *ompK* | 0.02026 |

**Table 10** Wrights inbreeding coefficient ($F_{ST}$), comparing Rhossili Bay (n=56) and Worms Head causeway (n=108) *Vibrio* isolates.

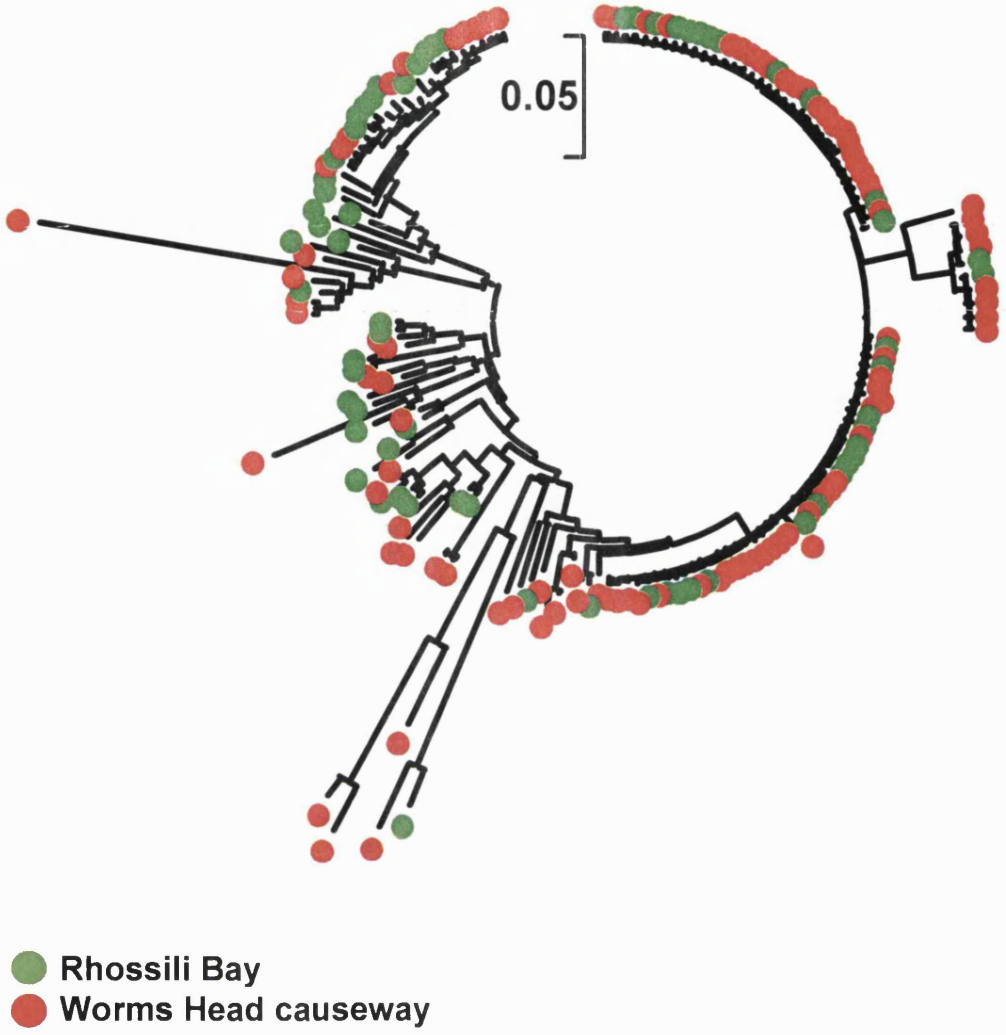



**Figure 13** Phylogenetic tree based on the Neighbour-Joining method, using 4 loci MLSA concatenated nucleotide sequences (n=164). Isolates are colour coded according to the sample site they were collected from.

5.3.9 Seasonal variation within the population

With seasonal variation already well documented within aquatic *Vibrio* populations (DePaola *et al.*, 2003), the four loci concatenated Neighbour-Joining tree (Fig. 14; Table 11) was coded according to sample collection date; S1=June 2004, S2=July 2004, S3=April 2005 and S4=August 2005. At first glance, we can see *Vibrio* isolates from the July 2004 sampling (S2) and August 2005 (S4) are distributed throughout the tree. Strains isolated during June 2004 (S1) and April 2005 (S3) are however observed at specific areas of the tree (June 2004 samples are located only in the large clade corresponding to *V. cyclitrophicus*, and April 2005 samples are located in a small region in the diverse centre region of the concatenated tree), suggesting some *Vibrio* species are abundant at different times of year. There are predominantly more isolates from the summer sampling dates, with only 4.9 % of isolates collected during spring. This lack of a representative sample for spring unfortunately does not allow an accurate analysis of seasonal variation within the UK *Vibrio* population (n=164), although comparisons based on sampling times are still informative by instead examining isolates for which *recA* species assignments are available (n=281).

| *Vibrio* species | S1 June 2004 | S2 July 2004 | S3 April 2005 | S4 August 2005 |
|---|---|---|---|---|
| *V. cyclitrophicus* | 36 | 50 | - | 11 |
| *V. tasmaniensis* | - | 20 | - | 5 |
| *V. kanaloae* | - | 28 | - | 6 |
| *V. splendidus* | - | 26 | 45 | 3 |
| *V. lentus* | - | - | - | 2 |
| *V. pomeroyi* | - | 6 | - | 1 |
| *V. harveyi* | 3 | 4 | - | - |
| *V. diabolicus* | 2 | 1 | - | 24 |
| *V. ichthyoenteri* | - | - | - | 4 |
| *V. pacinni* | - | 1 | - | - |
| *V. tubiashii* | - | - | - | 1 |
| *P. eurosenbergii* | - | - | 1 | - |

**Table 11** Number of *Vibrio* spp. recovered from each sampling date (n=281).

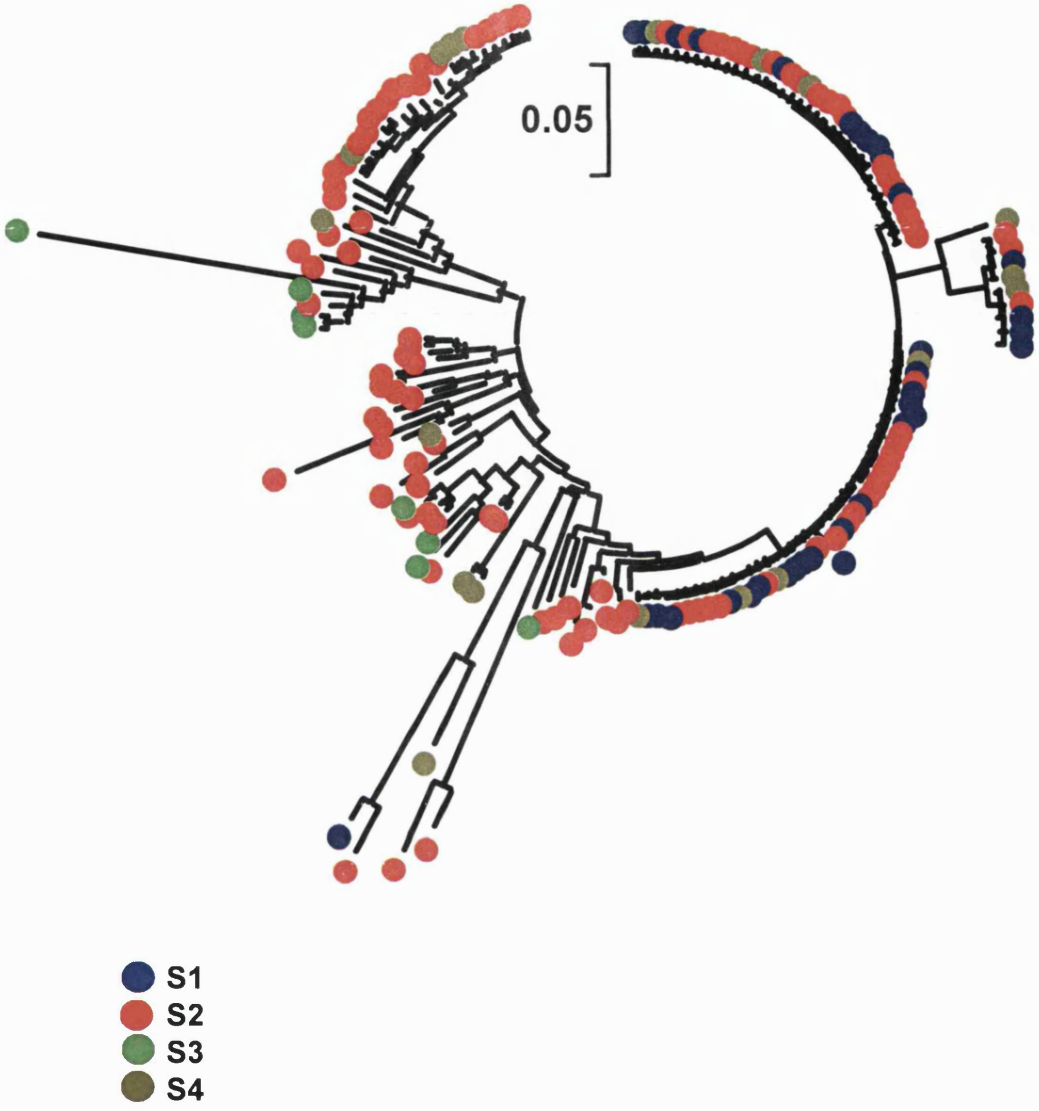



**Figure 14** Phylogenetic tree based on the Neighbour-Joining method, using 4 loci MLSA concatenated nucleotide sequences (n=164). Isolates are colour coded according to sample collection date; S1 = June 2004; S2 = July 2004; S3 = April 2005; S4 = August 2005.

The percentage of anemones from which *Vibrio* cultures were recovered ranges from 81.3 % (n=64) to 100 % (n=2), with no obvious differences apparent in overall *Vibrio* abundance in anemones (Chapter 2 Table 2). The *Vibrio* population is most abundant during late summer, with a total of eight *Vibrio* species isolated from July 2004 and nine from August 2005 (Table 11). Only six species were common to both late summer sampling dates; *V. cyclitrophicus, V. tasmaniensis, V. kanaloae, V. splendidus, V. pomeroyi* and *V. diabolicus*. The early summer sampling (June 2004) yielded three species, including *V. cyclitrophicus*, making it the dominant *Vibrio* species present during June 2004 and the summer months. As mentioned previously, the spring sampling yielded the fewest number of isolates within the MLSA study (n=8), and is represented by only one *Vibrio* species, *V. splendidus*. However, when all isolates for which *recA* species assignments are available are taken into account (n=281), *V. splendidus* is the dominant species present, representing 98 % of the samples collected at this time. *V. pacinni, V. tubiashii* and *P. eurosenbergii* are each represented by a single isolate recovered from July 2004, April 2005 and August 2005 respectively.

The lack of species diversity during the spring sampling suggests that *V. cyclitrophicus, V. tasmaniensis, V. kanaloae, V. pomeroyi, V. harveyi, V. diabolicus, V. ichthyoenteri, V. pacinni, V. tubiashii* and *V. lentus* may be in the VBNC state during this time, and it is only when the sea water temperature rises that they become abundant in UK waters. The SST was measured on each collection date (Table 1), indicating a variation of approximately 8°C between spring and summer. *V. splendidus* appears to be adaptable to seasonal water temperatures, being present during both spring and summer months. Further sampling is required during winter to identify which, if any, *Vibrio* species dominate at this time.

### 5.3.10 Effect of host association and sampling time on *Vibrio* population

An analysis of host association suggests that a wider variety of *Vibrio* species are able to obtain a reasonably high frequency within sea anemones than other eukaryotic hosts and sea water. Analysis of sampling time suggests that *V. cyclitrophicus* was most abundant during the summer months of 2004 and then declined in subsequent sampling. Unfortunately this presents a cause and effect

problem, as the July 2004 samples were biased towards sea anemones than during June 2004. Therefore, the dominance of *V. cyclitrophicus* in the June 2004 sample could be explained either by the wider range of hosts sampled or by temporal fluctuations in the population as a whole. Of these, the latter is more likely, as it would be expected that sampling a wider range of hosts would increase, rather than decrease *Vibrio* diversity. Because the temporal fluctuations within the *Vibrio* population as a whole appear extremely strong, and extremely rapid, it is difficult to find strong evidence for population association with sea anemones.

### 5.3.11 Clonal diversification is observed within the population

The differences between the various sampling times suggest that clones rapidly rise and fall in frequency within the population. In particular, the population was dominated by the *V. cyclitrophicus* clone in the June 2004 and July 2004 samples. Allelic profiles of 164 strains were assigned on the basis of nucleotide sequences for each gene for each of the four MLSA loci, and the structure of the population was examined using eBURST (Feil *et al.*, 2004). One hundred and twelve unique genotypes (STs) were determined, with eighty-eight of these STs represented by only a single isolate, while nine are represented by at least four isolates. The largest and most common clone was ST7, which accounts for nine isolates (5.4 % of all isolates). From a "snapshot" of the population (group definition 0/4 alleles in common) one major clonal complex is evident, which corresponds to the *V. cyclitrophicus* clade. This complex corresponds to a simple model of clonal variance. The one hundred and twelve STs were divided by eBURST into one major clonal complex, four doublets and 63 singletons (Fig. 15). The founder clone, ST7, supported by 88% bootstrap, corresponds to *V. cyclitrophicus* and conforms to a simple model of radial diversification (Fig. 16).

The identification of a single large clonal complex corresponding to *V. cyclitrophicus* (cc7) means it is possible to estimate whether SLVs and DLVs have arisen from their respective clonal ancestor by recombination or by point mutation (Feil *et al.*, 2003). The clonal complex shown in Figure 17 provides a total of 38 SLVs and 13 DLVs for analysis (Tables 12 & 13 respectively). 36.8 % of SLVs originate from a single nucleotide substitution, of which 42.9 % (n=6) of these allelic changes are unique,

consistent with them originating by de novo mutation. The remaining 57.1 % (n=8) of single nucleotide changes are not unique, therefore suggesting these STs have arisen by recombination. 36.8 % of the single locus variants involved 16 or more nucleotide changes, representing approximately 4 % divergence which is consistent with recombination occurring between named species. As with other MLST schemes, for example *Staphylococcus aureus* it would be expected that a model conforming to clonal diversification would predominantly have arisen by point mutation (Feil *et al.*, 2003). However this is not the case within the *Vibrio* population, and instead clonal diversification is driven predominantly by recombination, with approximately 5 times as many recombination events than point mutations (5.3:1) with SLVs. With DLVs this ratio is 11.5:1, however, the use of DLVs in determining levels of recombination should be treated with caution as we have less confidence that the patterns of decent have been correctly identified.

From Figure 18, *mdh* appears to be relatively conserved, with only 3 separate nucleotide changes observed within the population, two of which have arisen by point mutation. With 10 separate events (43.5 %), *recA* is the most common variant allele when less than 8 nucleotide changes occur. Furthermore, of the total number of single base pair changes, *recA* again dominates with 5 recombination events and one point mutation. Most recombination events, in particular those involving very diverse alleles, are observed within *recG* and *ompK*. These genes are the most diverse of the four loci, with large (greater than 16 bp) recombination events having taken place with both SLVs and DLVs. This is rather surprising with respect to *recG*, due to its essential role in Holliday junction migration, therefore would be expected to be relatively conserved and not involved in large scale recombination events. *ompK* encodes the receptor for Vibriophage KVP40 and is under diversifying selective pressures as a result of its interaction with phage within the environment, therefore, is under diversifying selection. With the largest recombination event totalling 78 bp (13.5 % of the region sequenced) in *ompK*, there is clearly a significant amount of gene flow between *Vibrio* species (average bacterial diversity is 2 %), in addition to gene replacements rather than point mutation as the method of change. As a result of this, it is hypothesised that the large scale recombination events occur between *Vibrio* species, and smaller sized recombination events are likely to occur within named species (Fig. 19). Within

SLVs, there are a total of 23 "within" recombination events and 15 "between" recombination events, meaning recombination is only marginally more likely (1.5 fold) within a *Vibrio* species as compared to between different species.

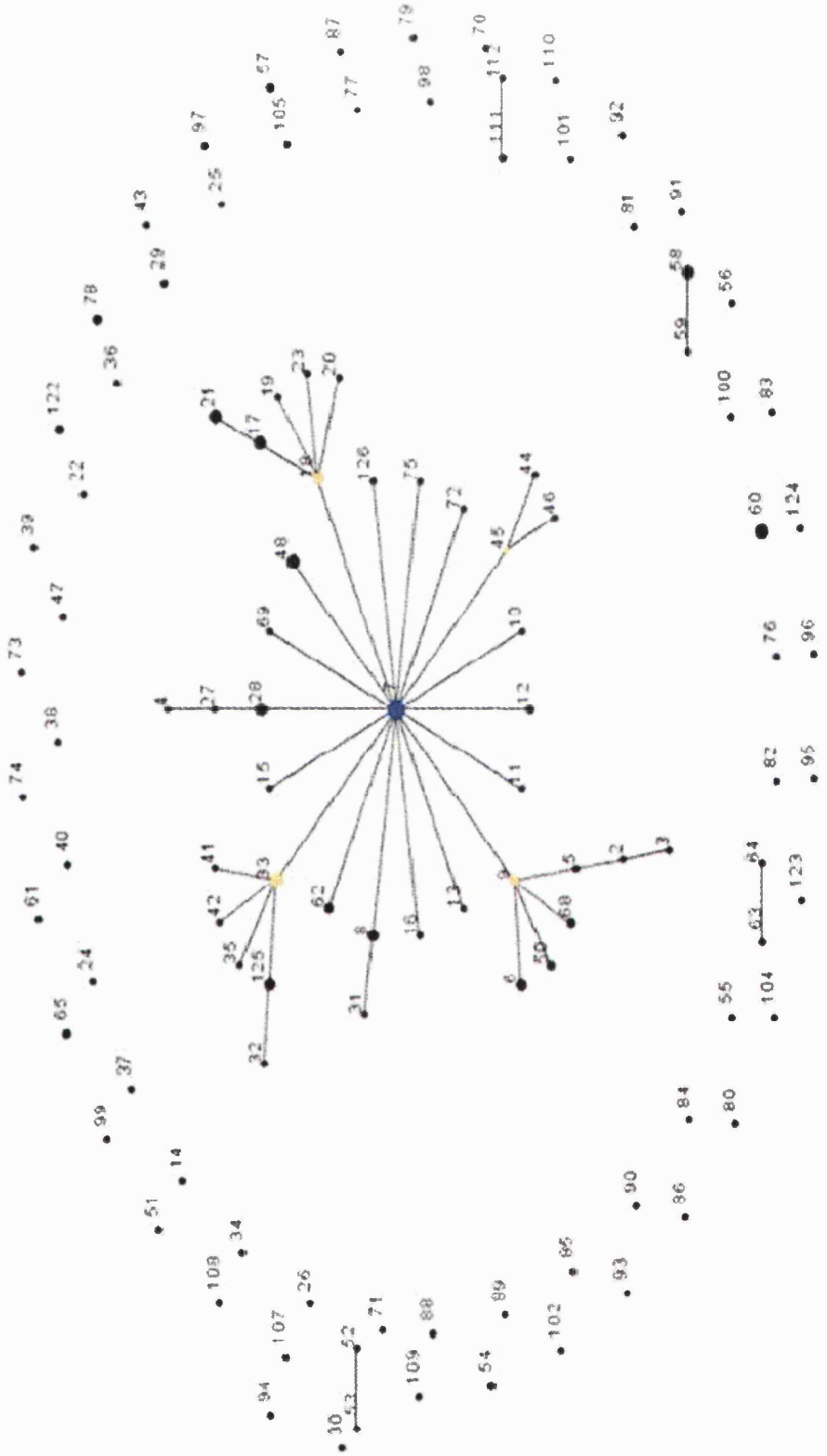


**Figure 15** A population "snapshot" showing the clusters of linked STs and unlinked STs for 164 *Vibrio* spp. (0/4 alleles in common). The predicted clonal ancestor is shown in blue, the subgroup founders shown in yellow. SLVs are shown as black lines. The sizes of the circles that represent each ST indicate their prevalence within the population. The numbers in the eBURST diagram correspond to each of the STs.

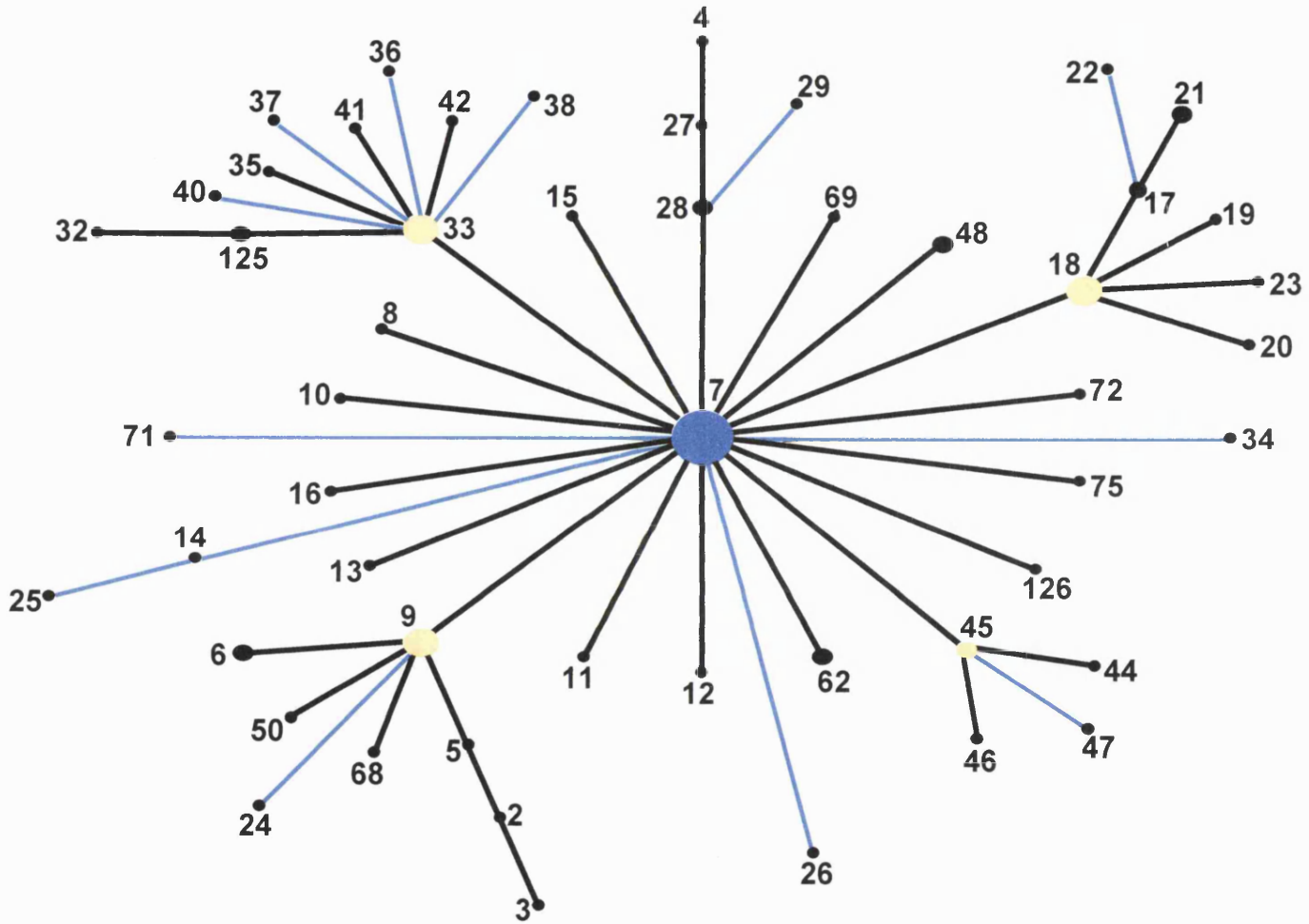



**Figure 16** Identification of a single clonal complex (cc7). The predicted clonal ancestor is ST7 and shown in blue. The clonal complex contains four subgroup founders shown in yellow. SLVs are shown as black lines and DLVs as light blue lines. This burst-o-gram has 2/4 alleles in common.

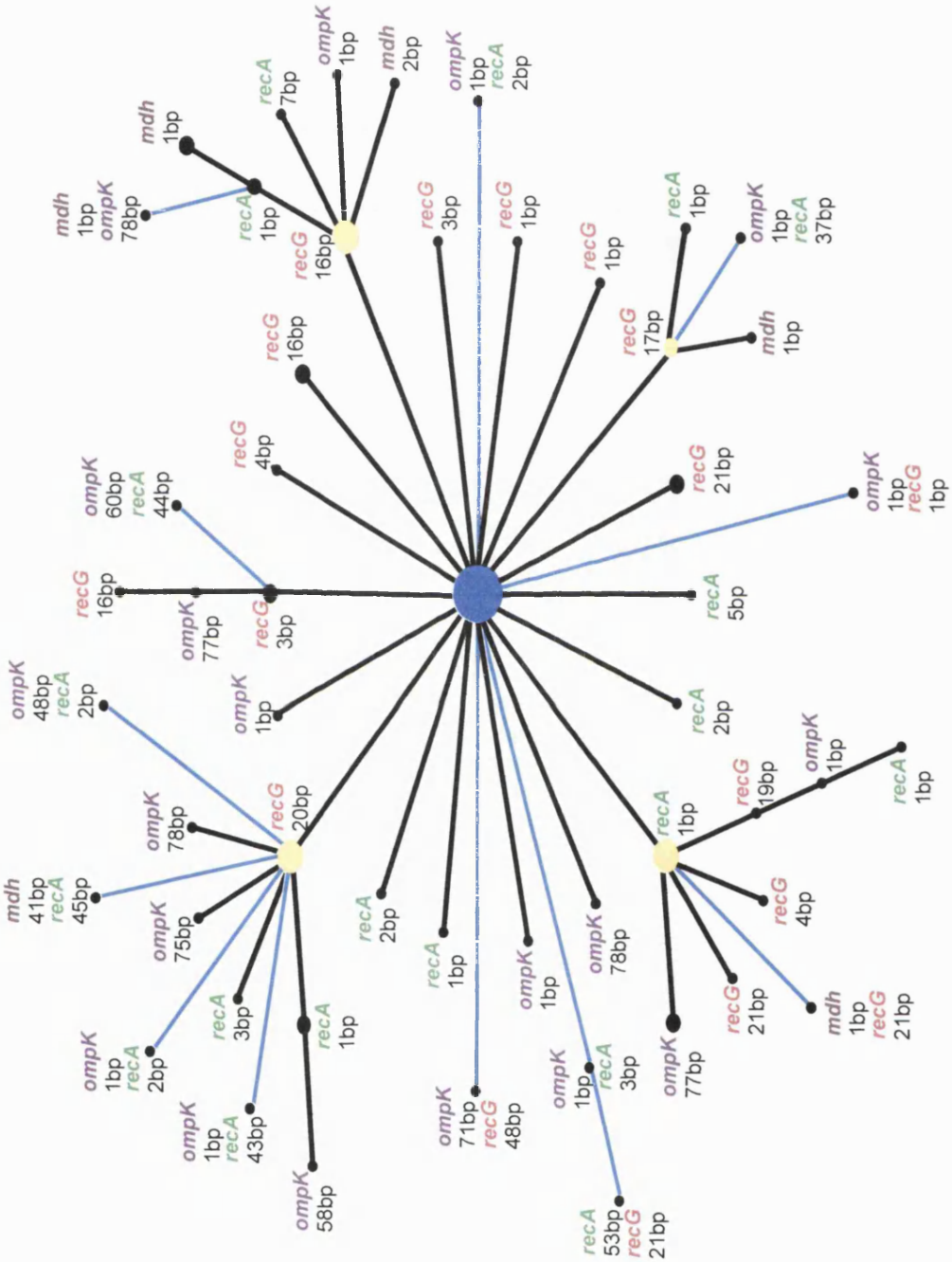



**Figure 17** Variant alleles within the SLVs and DLVs found within the clonal complex. Each SLV & DLV has beside it the variant loci in which change is detected and the number of nucleotide differences at that occur

| ST of clonal ancestor | ST of SLV | Variant locus in SLV | Ancestral allele | SLV allele | No. of bp differences | Mode of nucleotide change |
|---|---|---|---|---|---|---|
| 7 | 28 | *recG* | 4 | 8 | 3 | recombination |
| 7 | 69 | *recG* | 4 | 32 | 4 | recombination |
| 7 | 48 | *recG* | 4 | 16 | 16 | recombination |
| 7 | 18 | *recG* | 4 | 5 | 16 | recombination |
| 7 | 72 | *recG* | 4 | 35 | 3 | recombination |
| 7 | 75 | *recG* | 4 | 37 | 1 | mutation |
| 7 | 126 | *recG* | 4 | 12 | 1 | mutation |
| 7 | 45 | *recG* | 4 | 15 | 17 | recombination |
| 7 | 62 | *recG* | 4 | 26 | 21 | recombination |
| 7 | 12 | *recA* | 3 | 5 | 5 | recombination |
| 7 | 11 | *recA* | 3 | 8 | 2 | recombination |
| 7 | 9 | *recA* | 3 | 2 | 1 | recombination |
| 7 | 13 | *ompK* | 5 | 7 | 78 | recombination |
| 7 | 16 | *ompK* | 5 | 82 | 1 | mutation |
| 7 | 10 | *recA* | 3 | 6 | 1 | recombination |
| 7 | 8 | *recA* | 3 | 4 | 2 | recombination |
| 7 | 33 | *recG* | 4 | 10 | 20 | recombination |
| 7 | 15 | *ompK* | 5 | 80 | 1 | recombination |
| 28 | 27 | *ompK* | 5 | 4 | 77 | recombination |
| 27 | 4 | *recG* | 8 | 2 | 16 | recombination |
| 18 | 17 | *recA* | 3 | 2 | 1 | recombination |
| 18 | 19 | *recA* | 3 | 7 | 7 | recombination |
| 18 | 23 | *ompK* | 5 | 80 | 1 | recombination |
| 18 | 20 | *mdh* | 2 | 5 | 2 | recombination |
| 17 | 21 | *mdh* | 2 | 21 | 1 | mutation |
| 45 | 44 | *recA* | 3 | 2 | 1 | mutation |
| 45 | 46 | *mdh* | 2 | 6 | 1 | mutation |
| 9 | 5 | *recG* | 4 | 2 | 19 | recombination |
| 9 | 68 | *recG* | 4 | 32 | 4 | recombination |
| 9 | 50 | *recG* | 4 | 18 | 21 | recombination |
| 9 | 6 | *ompK* | 5 | 4 | 77 | recombination |
| 5 | 2 | *ompK* | 5 | 2 | 1 | recombination |
| 2 | 3 | *recA* | 2 | 4 | 1 | recombination |
| 33 | 125 | *recA* | 3 | 2 | 1 | recombination |
| 33 | 35 | *recA* | 3 | 6 | 3 | recombination |
| 33 | 41 | *ompK* | 5 | 73 | 75 | recombination |
| 33 | 42 | *ompK* | 5 | 86 | 78 | recombination |
| 125 | 32 | *ompK* | 5 | 4 | 58 | recombination |

**Table 12** Variant alleles within the SLVs found within the *V. cyclitrophicus* clonal complex.

| ST of clonal ancestor | ST of DLV | Variant locus in DLV | Ancestral alleles | DLV alleles | No. of bp differences | Mode of nucleotide change |
|---|---|---|---|---|---|---|
| 7 | 34 | *ompK* | 5 | 6 | 1 | recombination |
| | | *recA* | 3 | 36 | 2 | recombination |
| 7 | 26 | *ompK* | 5 | 6 | 1 | recombination |
| | | *recG* | 4 | 7 | 1 | mutation |
| 7 | 14 | *ompK* | 5 | 42 | 1 | recombination |
| | | *recA* | 3 | 47 | 3 | recombination |
| 7 | 71 | *ompK* | 5 | 29 | 71 | recombination |
| | | *recG* | 4 | 34 | 48 | recombination |
| 14 | 25 | *recA* | 47 | 20 | 53 | recombination |
| | | *recG* | 4 | 6 | 21 | recombination |
| 45 | 47 | *ompK* | 5 | 6 | 1 | recombination |
| | | *recA* | 3 | 41 | 37 | recombination |
| 9 | 24 | *mdh* | 2 | 3 | 1 | recombination |
| | | *recG* | 4 | 6 | 21 | recombination |
| 33 | 40 | *ompK* | 5 | 57 | 1 | recombination |
| | | *recA* | 3 | 26 | 43 | recombination |
| 33 | 37 | *ompK* | 5 | 6 | 1 | mutation |
| | | *recA* | 3 | 4 | 2 | recombination |
| 33 | 36 | *mdh* | 2 | 25 | 41 | recombination |
| | | *recA* | 3 | 48 | 45 | recombination |
| 33 | 38 | *ompK* | 5 | 25 | 48 | recombination |
| | | *recA* | 3 | 21 | 2 | recombination |
| 28 | 29 | *ompK* | 5 | 52 | 60 | recombination |
| | | *recA* | 3 | 24 | 29 | recombination |
| 17 | 22 | *mdh* | 2 | 3 | 1 | recombination |
| | | *ompK* | 5 | 7 | 78 | recombination |

**Table 13** Variant alleles within the DLVs found within the *V. cyclitrophicus* clonal complex.

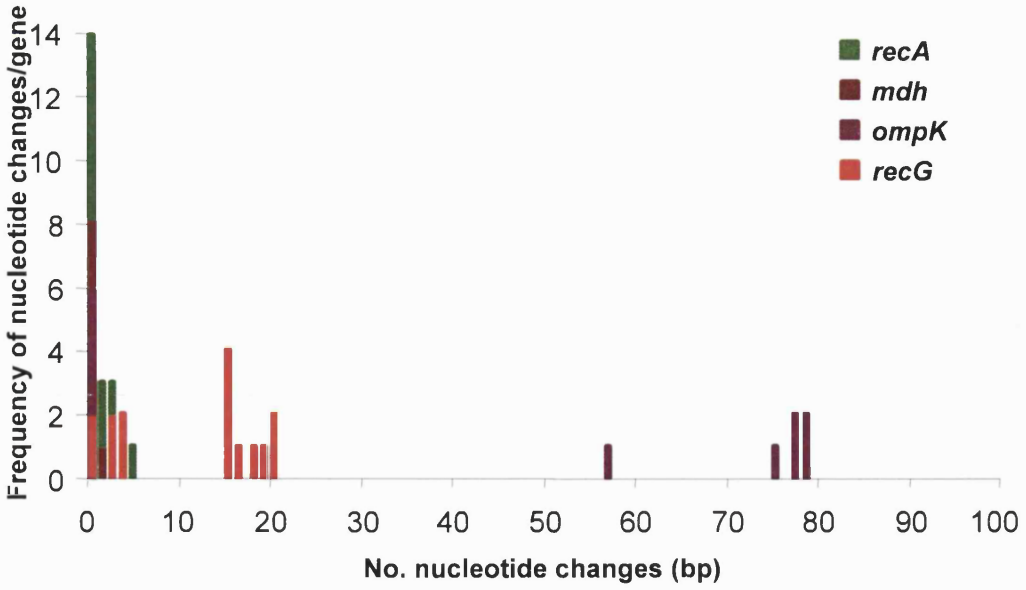



**Figure 18** Frequency of variant loci observed from SLVs. The four MLSA loci have each been colour coded.

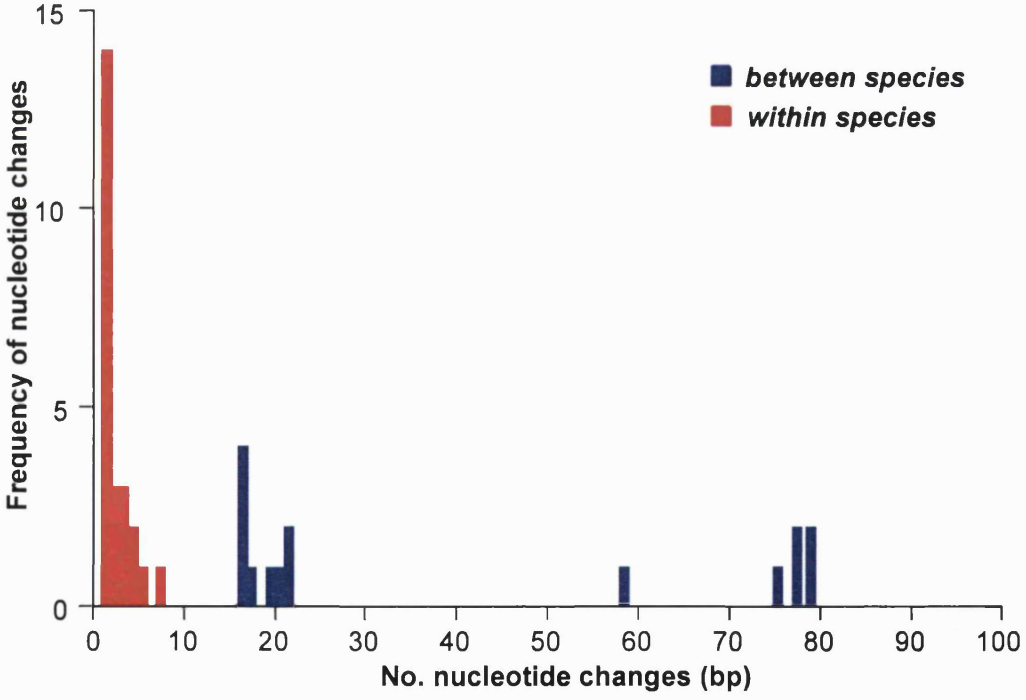



**Figure 19** Distribution of nucleotide changes observed from SLVs.

With high levels of recombination and high dS/dN ratios, amino acid allelic profiles were entered into eBURST to obtain a clearer picture of the UK *Vibrio* population dynamics (group definition, 0/4 alleles in common). As there are only four loci available, and a high level of diversity within the sample, the nucleotide sequence alleles have limited power to infer pathways beyond clonal founders and their immediate descendents. To adjust for this, the alleles were translated, thus "slowing down" the rate of allelic change and providing a broader view of the population structure. Four clonal complexes are observed (Fig. 20), aa-cc2, aa-cc9, aa-cc15, aa-cc45, each conforming to a model of radial diversification. aa-cc15 corresponds to the large clonal complex observed in Figure 16 and using *recA* species assignments, the STs correspond predominantly to *V. cyclitrophicus* and also to *V. splendidus* and *V. kanaloae*. aa-cc2 is also a mixed species clonal complex consisting of *V. tasmaniensis* and *V. kanaloae*. The SLVs of both aa-cc9 and aa-cc45 form single species clonal complexes; aa-cc9 consists of V. *kanaloae* and aa-cc45 of *V. splendidus*. Of the variant alleles presented at SLVs, *recA* and *recG* occur at the same frequency (20.5 %), with *mdh* occurring least often (9 %) and *ompK* as expected the variant allele in 50 % of STs (Fig. 21). This is consistent with data from nucleotide eBURST. Although there is clearly a large amount of recombination taking place within the genus, these changes are dominantly synonymous in *recA, recG* and *mdh*, indicating the genus is under considerable purifying selection.

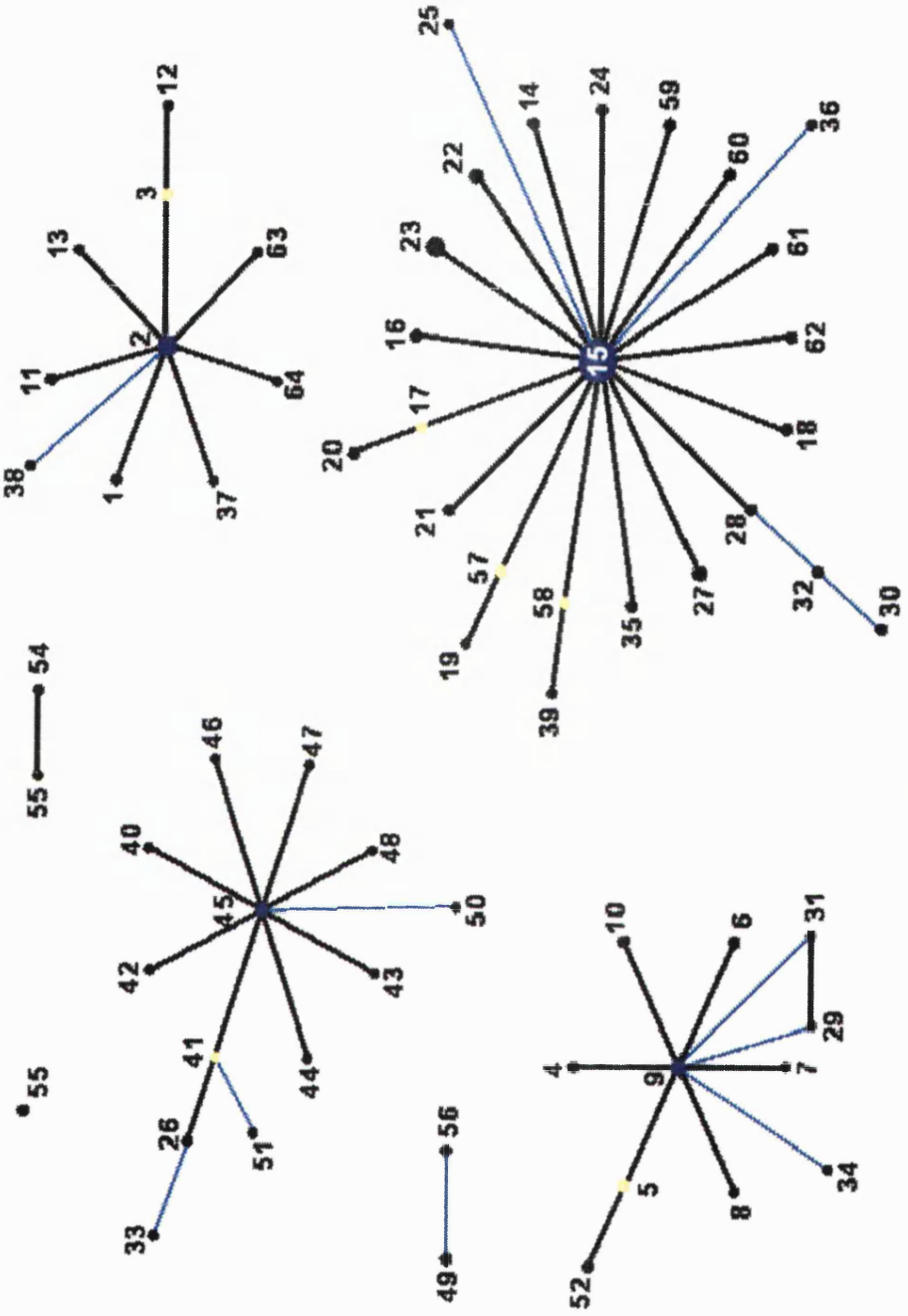



**Figure 20** A population "snapshot" showing the clusters of linked STs and unlinked STs for 164 *Vibrio* spp. viewed at amino acid level (0/4 alleles in common). Four clonal complexes, aa-cc9, aa-cc45, aa-cc2 and aa-cc15, are observed each with sub-group founders. aa-eBurst provides a clearer overview of the population structure than with na-eBurst.

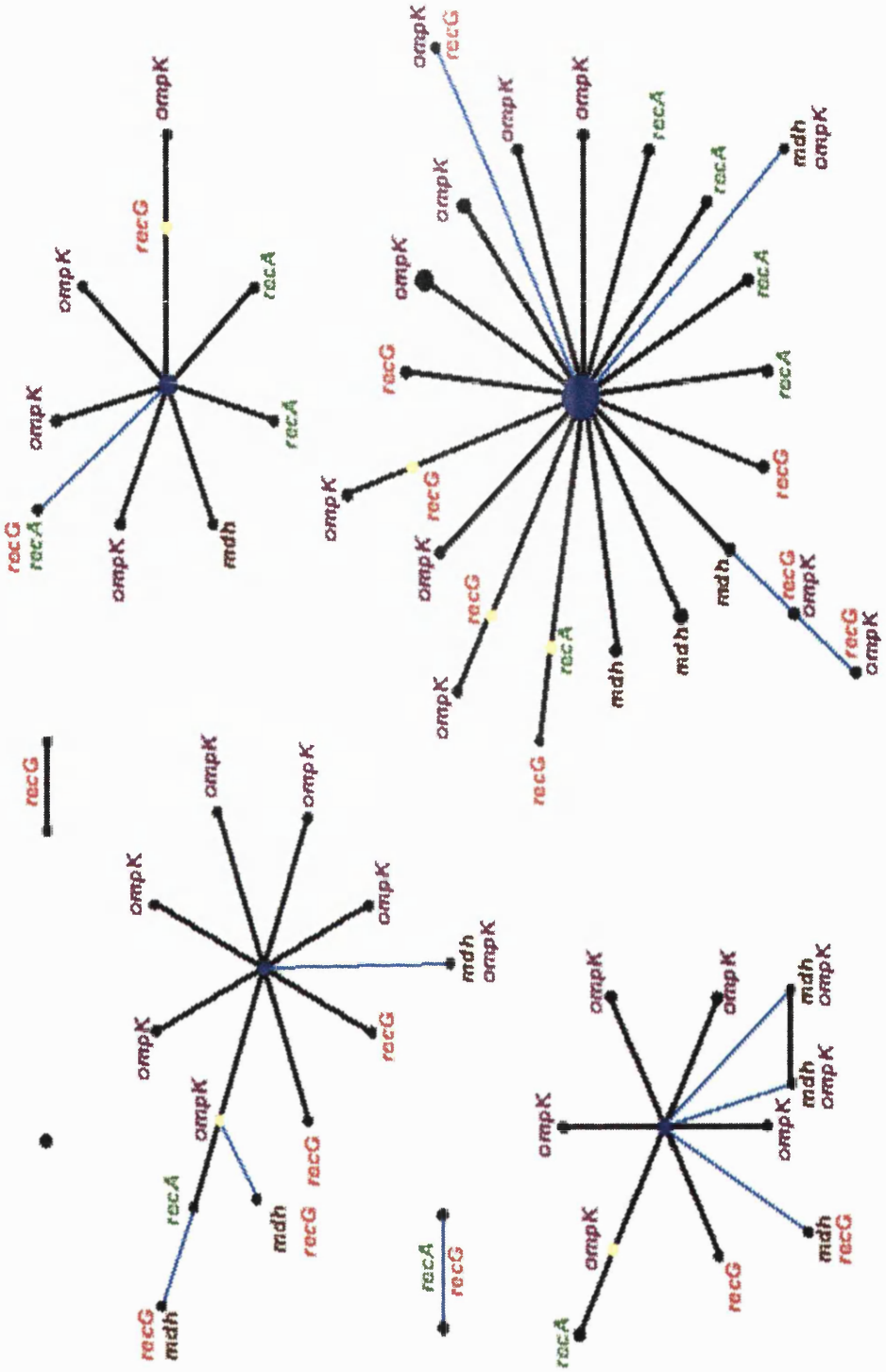



**Figure 21** Variant alleles within the SLVs and DLVs found within each of the amino acid clonal complexes. Each SLV & DLV has beside it the variant loci in which change is detected.

## 5.4 DISCUSSION

The sequencing of complete genomes has confirmed the high frequency of genetic exchange in prokaryotes (Ochman *et al.*, 2000) and of gene duplications in eukaryotes (Lynch and Conery, 2000). It is therefore potentially hazardous to infer species phylogeny from single gene trees. Like most bacterial taxa, *Vibrio* species have traditionally been assigned on the basis of 16S rRNA sequence or biochemical criteria. With the ever increasing pace and popularity of multi-locus sequencing, 16S rRNA has been proven to have poor discriminatory power when assigning species; *V. alginolyticus, V. parahaemolyticus, V. harveyi, V. campbellii, V. natriegens and V. rotiferianus* are virtually indistinguishable on the basis of the 16S rRNA sequence (Thompson *et al.*, 2004b). Furthermore, the use of only a single locus increases the vulnerability of the data to phylogenetic distortion by recombination. Although 16S rRNA is not believed to recombine freely, notable exceptions have been reported in a number of taxa. For example, Schouls *et al.* reported that members of the *Streptococcus anginosus* group (SAG), which is separated into three distinct *Streptococcus* species, display unusual sequence diversity that is most likely the result of gene transfer and recombination.

Alternative phylogenetic markers have long been sought after, with genus *Vibrio* being no exception. Recent studies have focused on *recA* as a suitable candidate gene. *recA* was first used as an aid to bacterial classification in 1993, when Lloyd and Sharp compared twenty-five bacterial *recA* sequences, concluding that there was a broad concordance between *recA* and 16S rRNA sequences (Lloyd and Sharp, 1993). The four MLSA genes used within this study suggest there are in fact two genes that differentiate well between *Vibrio* species. *recA* is the most evolutionary conserved of the loci examined, with the highest synonymous substitution rate and the lowest nucleotide variation, it is under the strongest purifying selection making it a good candidate gene for species identification. This is consistent with previous phylogenetic studies (Lloyd and Sharp, 1993; Stine *et al.*, 2000; Thompson *et al.*, 2004a), thus making it a good candidate as an alternative phylogenetic marker to 16S rRNA. However, *recA* is subject to frequent recombination events, as will be discussed in Chapter 6, therefore is should not be solely used for the identification of *Vibrio* spp. *mdh* could also be considered a useful molecular chronometer, due to

its high level of congruence with the species assignments on the *recA* tree (Figure 8), and its modest dS/dN levels indicating it is also under purifying selection. *recG* and *ompK* should not be considered as suitable alternatives to 16S rRNA; firstly their individual phylogenies do not differentiate well between all *Vibrio* species; within their respective gene segments, they have 9.4 % and 15.7 % nucleotide diversity respectively; both genes have high levels of variable sites (over 50 % of the region of gene sequenced is variable in both genes); *ompK* has the lowest dS/dN ratio of 3.73 is observed with *ompK*, therefore it has the weakest purifying selection of all the genes, bordering on the brink of positive selection. *ompK* encodes for the broad host range vibriophage receptor KVP40, therefore is in constant exchange of genetic material from the environment. This gene is therefore undergoing rapid evolution and diversifying selection.

With the realisation that individual genes fail to accurately resolve named species, more emphasis has been placed upon multi-locus approaches (Hanage *et al.*, 2005). By concatenating sequence data, the multiple loci buffer against the distorting effects of inter-species recombination. Of course it is essential to ensure suitable loci are chosen for the analysis, in particular those that are not under positive selection, otherwise the effects of recombination will prove to be too strong, thus distorting the tree. Concatenation of the four protein coding genes results in partially distorted species clusters, which become more resolved when *ompK* is removed from the concatenated data. This gene has been included in this MLSA study, due to the vital information it will provide concerning global adaptation (this will be discussed further in Chapter 6).

As marine bacteria, *Vibrio* species understandably are affected by environmental factors, such as salinity, temperature and in some cases the abundance of host organisms (Lee and Ruby, 1994). Seasonal variation can result in fluctuations in *Vibrio* populations, in particular with a change in temperature (DePaola *et al.*, 2003). *Vibrio* populations are directly correlated with water temperature, therefore a decline in numbers is observed in winter months. Wright *et al.*, (1996) reported that within the *V. vulnificus* population in Chesapeake Bay, bacterial levels on average do not increase until May when the water temperature is significantly warmer; between February and April, no *V. vulnificus* population is detectable. This is consistent with

the data collected from the UK site, in which *V. cyclitrophicus, V. tasmaniensis, V. kanaloae, V. pomeroyi, V. harveyi* and *V. lentus* are only observed during summer and not in spring. However, *V. splendidus* is detectable in both spring and summer, suggesting this species is perhaps more "hardy" than the others and is able to regenerate from the VBNC state at colder temperatures. To ascertain exactly what is happening to the *Vibrio* population during colder temperatures, sampling would need to be performed in winter months, between November and February, and also during spring. Most VBNC studies have focused on *Vibrio* species which are human pathogens, using techniques such as PCR (Brauns *et al.*, 1991), fluorescent monoclonal antibody detection (Brayton *et al.*, 1987), and immunoassay test kits, e.g. CholeraSMART (New Horizon Diagnostics, Columbia, USA) which requires no culturing of cells (Colwell, 2000). With the average UK winter sea water temperature between 8 – 12°C we can assume that the majority of the summer *Vibrio* species (*V. cyclitrophicus, V. tasmaniensis, V. kanaloae, V. pomeroyi, V. harveyi* and *V. lentus*) would be in a VBNC state in winter, therefore traditional detection methods would be redundant.

In addition to temperature, the complex relationship between phytoplankton and *Vibrio* bacteria also contributes to population structuring. It is well documented within Bangladesh, bimodal cholera outbreaks are common-place and are the result of increases in copepod density. These tiny zooplankton graze on phytoplankton which bloom and increase in number during spring and autumn. As the number of copepods increase, so do *Vibrio* numbers resulting in an epidemic. A single copepod can carry up to $10^4$ *V. cholerae* cells, and with the infective dose of cholera being $10^3$ CFU (Colwell, 1996), it is clear the devastating effects of phytoplankton blooms in areas with little or no water treatment systems (Colwell, 1996; Heidelberg *et al.*, 2002). Fortunately within the UK, cholera outbreaks are no longer considered a problem and scientists monitoring phytoplankton blooms around the coast of Britain do so for alternative reasons. The nearest documented phytoplankton blooms to the Gower, Wales are throughout the Bristol Channel (Cloern *et al.*, 1985). These blooms have been monitored during spring, a period in which little species diversity was observed within the UK *Vibrio* population, so perhaps *V. splendidus* is

the first of the *Vibrio* species to "wake up" from the VBNC state. Only with further sampling, can this hypothesis truly be tested.

Unfortunately, the salinity of the sea water was not tested at the sample sites, so no conclusions can be drawn as to its impact on the UK population.

*Vibrio* species interact with a vast range of eukaryotic hosts within aquaculture, with several species isolated from marine organisms in recent years (Lee and Ruby, 1994). *Vibrio* bacteria form a major constituent of bacterio-plankton and colonise a vast array of hosts, although almost all of these interactions have been studied due to their commercial or public health relevance, rather than from an evolutionary perspective. This study aims to provide more information as to the evolutionary path of an environmental *Vibrio* population. All of the seven *Vibrio* species isolated from the two sites were found colonising sea anemones. *V. cyclitrophicus* and *V. kanaloae* were isolated from both the sea water and periwinkle samples; *V. cyclitrophicus* and *V. splendidus* found colonising limpets; and finally crabs and mussels were colonised by only *V. cyclitrophicus*. We can therefore conclude, i) *V. cyclitrophicus* is the dominant and possibly most generalist species within the population, isolated from at least 5 eukaryotic hosts, ii) no single *Vibrio* species is specific to any one eukaryotic species (*V. cyclitrophicus, V. tasmaniensis, V. pomeroyi, V. harveyi, V. lentus*, *V. kanaloae* and *V. splendidus* all have the ability to colonise sea anemones), and iii) *V. tasmaniensis, V. lentus, V. pomeroyi* and *V. harveyi* were not recovered from sea water samples. In addition, more then one *Vibrio* species can colonise a eukaryotic host at one time. For example, strain S2aA29vi, identified as *V. pomeroyi* was isolated from sea anemone number 29, as was S2aA29ix identified as *V. kanaloae.* This shows a possible route by which recombination can take place. This reinforces the hypothesis that *Vibrio* species are able to colonise and successfully clonally diversify within host populations.

As discussed earlier, seasonal variation has a substantial effect on the *Vibrio* population. With *V. tasmaniensis, V. cyclitrophicus, V. pomeroyi, V. harveyi, V. kanaloae* and *V. lentus* only detectable during summer sampling, these species will be in direct competition with *V. splendidus* which is able to colonise their eukaryotic hosts (e.g. sea anemones and limpets) during spring and summer; thus giving them

an ecological advantage. With *V. kanaloae* colonising sea anemones, periwinkles and present in sea water, in addition to *V. tasmaniensis* colonising sea anemones and periwinkles, it is possible they hold an ecological advantage within a select niche, over the remaining species. Very little is known about both these species, apart from them being closely related by 16S rRNA and *recA* (Thompson *et al.*, 2004c). Within this environmental MLSA study, both these species have been observed in the same clusters on both the individual and concatenated trees. Further analysis, including additional MLSA loci, will provide more information on this close species association.

To assess the recent evolutionary events of the UK population, allelic profiles and genotypes were generated and entered into eBURST (Feil *et al.*, 2004). At nucleotide level, a large radially diversified clonal complex was observed corresponding to the large *V. cyclitrophicus* clade on the concatenated phylogenetic tree. To assess the overall population structure at the genus level, translated sequence was used for the generation of allelic profiles and sequence types. Within the UK *Vibrio* population, 4 clonal complexes are observed, each of which conform to radial diversification. Although there is clearly a large amount of recombination taking place within the genus, these changes are dominantly synonymous indicating the genus is under considerable purifying selection. As MLSA becomes increasingly used for genus wide analysis of environmental populations, this approach may prove very powerful in resolving inter-species relatedness.

The detection of the large *V. cyclitrophicus* clone was unexpected given the data of Thompson *et al.*, who sampled Vibrios within the water column of a coastal island in Massachusetts to determine the genotypic diversity and overall population size of the natural *Vibrio* population quantitatively using *V. splendidus* (Thompson *et al.*, 2005). Sequences were characterised using *hsp60* and rarely found the same genotype represented in more than one isolate; approximately 1000 distinct genotypes were observed. Thompson *et al.* interpreted this as evidence for very little clonal structure within the natural *Vibrio* population, whereas the data presented here finds a predominant clone. Possible explanations for this contradictory result include the use of an unusually variable locus by Thompson *et al.*, or the effect of sampling from

multiple eukaryotic hosts within this study; although it is clear that a great variety of different species can colonise sea anemones.

The detection of clonal founders and SLVs allows an estimate of the relative contribution of point mutation and recombination to clonal diversification using the method of Feil *et al.*, (1999). The estimate for the large *V. cyclitrophicus* clone is approximately 5:1 in favour of recombination, which is very similar for estimates of the recombinogenic human pathogenic species *N. meningitidis* and *S. pneumoniae*. Furthermore, inter-species recombination appears to be common, with clear mosaicism evident in the *recG* sequences. The *Vibrio* genus therefore appears to be similar in this regard to the *Neisseria* genus, where clusters corresponding to named species are evident on concatenated phylogenies, but recombination between species in single named species is frequently observed. Maynard Smith *et al.* described such a scenario as akin to a "commonwealth" of individual clusters; on the one hand distinct but also sharing DNA through rampant recombination (Smith *et al.*, 2000). Similarly, it seems likely that the "epidemic" model of clonal expansion is likely to be apt for *Vibrio* populations (Smith *et al.*, 1993). Such a population structure is characterised by frequent recombination but occasional rapid expansion of adaptive clones; in the case of *Vibrio* this may be due to abiotic seasonal (notably temperature) fluctuations or changes in the host populations.

# CHAPTER SIX

# THE GLOBAL BIOGEOGRAPHY AND EVOLUTION OF *VIBRIO* SPECIES

# 6.1 INTRODUCTION

## 6.1.1 Microbial biodiversity

The marine environment is of immense importance as a reservoir for bacterial biodiversity. Microbial cell concentrations are estimated to be $10^5$ cells $ml^{-1}$ in the ocean surface layer, and a recent study using random shotgun cloning methods have unveiled a seemingly bottomless well of bacterial molecular variation in the Sargasso Sea (Venter *et al.*, 2004). The quantification of this variation is only just beginning, so our understanding of the processes leading to the emergence and maintenance of this variation is very poor. As for other fundamental aspects of evolution, the majority of the theory and empirical evidence is based on eukaryotic systems, in particular relating to the importance of selective and stochastic process in shaping species distribution. For example, high levels of intra-species diversity, along with a broad geographical range, are thought to provide a kind of "insurance policy" in the face of environmental change, as low levels of genetic diversity or small isolated populations are more vulnerable to extinction. However, it is unclear at present to what extent our understanding of microevolution in eukaryotes can be extrapolated to microbial populations.

In Chapter 5, the variation of *Vibrio* species within a single localised population isolated from South Wales was considered. In this chapter, the analysis has been extended to consider the distribution of variation on a global scale by comparing the population isolated in South Wales, with that of two populations from North Island, New Zealand. This approach encompasses a range of geographical distances, the UK and NZ sites are approximately 11,500 km apart, whereas the two sites in NZ are separated by approximately 1000 km (by sea). Such comparisons over broad geographical scales are essential for a complete picture of the biogeography of a given species or genus.

## 6.1.2 Local and global diversity

Consider first community, or local diversity. This is in a large part determined by the degree of competition between species, which in turn reflects ecological factors such as predation, productivity and environmental heterogeneity. Interspecific

competition can influence species diversity due to dominant species outcompeting subordinate species in the local environment. This in itself is influenced by the productivity of a given environment; when productivity is low, the competition between species is stronger, because the amount of resources/nutrients shared by the species is more limited (Storch, 2001). In addition, extremely high levels of productivity may also promote competition, in which the less adapted species are outcompeted. Therefore species biodiversity is expected to be lower in cases of extremely low or extremely high productivity, in the latter case a few highly adapted species will exploit the majority of the resources, hence outcompete less well adapted forms (Martiny *et al.*, 2006).

A second biotic factor affecting local diversity is environmental heterogeneity and niche availability. A high degree of environmental heterogeneity promotes the coexistence of species, where the availability of multiple niches will minimise direct competition (Storch, 2001).

Global diversity is influenced by speciation and extinction dynamics driven primarily by abiotic factors such as large scale geographical and/or climate changes combined with limited migration. In cases where populations are separated by very large distances, it can be speculated that the relative importance of local adaptation in explaining population differences is lower than for more localised comparisons. This is because widely separated populations are expected to have diverged substantially by genetic drift and other processes, and these differences are likely to overwhelm the more subtle adaptations owing to differing habitats (Martiny *et al.*, 2006; Storch, 2001).

### 6.1.3 Framework for microbial biogeography

Although much is known about the biogeography of eukaryotes, far less data is available for bacteria, however patterns of variation and the diversification of natural bacterial populations on differing geographical and temporal scales is a rapidly growing area of research. The prevailing hypothesis for bacterial biogeography is based on the axiom of the Dutch microbiologists Baas-Becking and Beijerinck, who stated:

*"Everything is everywhere, but the environment selects"*

(Beijerinck, 1913).

This widely quoted statement contains two direct assumptions which require separate examination. "*Everything is everywhere...*" assumes that microbes are sufficiently small, and have a sufficiently large population size, that they are able to permeate all ecological and geographical boundaries which are known to constrain distribution ranges of eukaryotes (Martiny *et al.*, 2006). This enormous dispersal capability will lead to the rapid erasure of past evolutionary events (i.e. historic events), which in turn prevents spatial variation. A trivial consideration of the ecological specialisations of prokaryotic genera suggests that the "*Everything is everywhere...*" hypothesis can only hold within certain phylogenetic constraints; soil or marine bacteria are rarely recovered from the mammalian gut (an exception to this case is *V. cholerae*) and obligate endosymbionts, such as *Wolbachia*, are by definition, never observed outside their host. However, might the assumption *"Everything is everywhere... "* hold within certain phylogenetic constraints? For example, might a given Order be everywhere? A given Genus? A given multi-locus genotype? A given allele? What exactly is it that we can expect to find everywhere? One way by which we can simplify this issue is by first identifying a truly "cosmopolitan" species (one with a global distribution), and then by asking whether each "strain" of such a species shares the global distribution of the species as a whole.

The best evidence currently available for addressing this question is for microbial species associated with humans, which can occasionally cause serious infection. Early evidence for human associated *E. coli* based on MLEE data suggested that the vast majority of global diversity can be detected at a local level (Whittam *et al.*, 1983). Similarly, many, if not most of the major clinically relevant clones of *Staphylococcus aureus* or *Neisseria meningitidis* have achieved a global distribution. In these cases, the frequent migration of asymptomatically colonised individuals (in particular since the advent of air travel), and frequent human-to-human transmission can account for an apparent loss of geographical structure. However, studies on these populations are typically motivated by their clinical relevance rather than from

a population biology perspective, hence the evidence is limited by the absence of a representative sample of the population. All too often, extensive collections of clinical isolates exist for most pathogens, with little or no interest in the collection of samples from asymptomatic carriage/environmental organisms. This sampling bias undermines the principle understanding of local adaptation on a global scale and urgently needs to be addressed.

The second part of this quote "*...the environment selects*" refers to local adaptation and environmental heterogeneity. Together, the statement implies that biogeographical patterns are not concerned with the absolute presence or absence of different forms, merely their relative abundance. This remains a useful standpoint, as it is basically impossible to prove the complete absence of any bacterial form from a particular habitat. In a recent review, Martiny *et al.*, (2006) discussed the extent to which microbial biogeography is likely to differ from our current understanding for eukaryotes. They distinguished two distinct processes which could lead to the non-random geographic distribution of microbial species or genotypes: local adaptation and historical contingencies. If local adaptation (habitat) is of primary importance, then the correlation between diversity and geographic scale should be weak. Conversely, if local adaptation is less important then this correlation should be strong, even if the samples are recovered from very different habitats on a local scale. They suggest that both factors contribute to the contemporary distribution patterns, but the relative importance of "habitat" over "province" (geographic region) decreases with increasing geographic scale (Martiny *et al.*, 2006).

As with biodiversity, the basic unit of biogeography is the species. Within the eukaryotic kingdom the "species" concept is well established by biological boundaries and is underpinned by evolutionary and ecological processes. However, within prokaryotes species demarcation is somewhat arbitrary and not as clearly defined. For example, pathogenic bacteria are typically classified according to the disease they cause regardless of other ecological or evolutionary considerations (Gevers *et al.*, 2005). This discrepancy could result in inconsistencies within bacterial systematics, with knock-on implications for biogeographical studies.

The incorrect use of specific microbial species gene markers further complicates our understanding of microbial biogeography. As mentioned previously in Chapter 1, the most common methods used to define prokaryotes species are DNA-DNA hybridisation and rRNA gene sequencing. Most previous studies have focused on sequencing 16S rRNA, which due to the conserved nature of the gene, does not yield sufficient resolution for species identification. A recent study by Thompson *et al.*, analysed the usefulness of *recA* gene sequences as an alternative phylogenetic and/or identification marker for *Vibrio* bacteria. Significant differences in similarity between 16S rRNA and *recA* (of the 16S rRNA similarity values above 98 %, the *recA* similarities ranged from 83 % to 99 %) were observed, indicating that *recA* was much more discriminatory then 16S rRNA as a phylogenetic marker for *Vibrio* species (Thompson *et al.*, 2004a). Therefore, it is essential that the correct choice of species gene marker is made to aid, not hinder, biogeographical studies.

### 6.1.4 Species dispersal in bacteria

One of the questions raised by the *"Everything is everywhere..."* hypothesis is what are the mechanisms behind microbial species dispersal and colonisation (Martiny *et al.*, 2006)? Two important mechanisms which shape microbial biogeography are active dispersal (e.g., an organism dispersing itself by active propulsion, such as propelling itself micrometers with its flagella) and passive dispersal (e.g., being carried thousands of kilometres by wind or ocean currents), however, these both have limiting factors. Active dispersal is proportionally related to the size (body mass) of an organism (Fig. 1a), therefore microorganisms have a very limited capacity to cross significant geographical boundaries under their own propulsion (Martiny *et al.*, 2006). They can however, travel great distances over the course of many generations by this active method, although it is likely some genetic divergence would have occurred over time from the primary population. Passive dispersal on the other hand is not limited by body mass (Fig. 1b), with microorganisms dispersed globally or over much shorter distances, creating non-random distributions of microbial assemblages (Martiny *et al.*, 2006). A major limiting factor with passive dispersal is whether a microorganism is able to propagate and outcompete the new local population it will encounter which may be better adapted to the new ecological conditions.

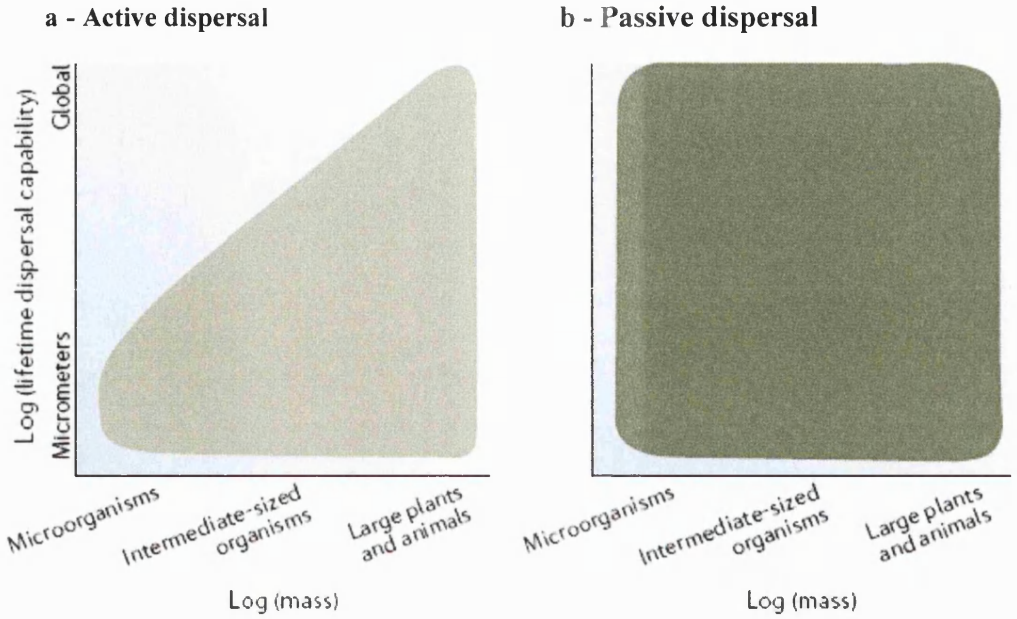



**Figure 1** Hypothetical relationship between body mass (at an organism's largest life-stage) and lifetime dispersal capability. The relationship is dependent on either a) active or b) passive dispersal. Figure from Martiny *et al.*, (2006).

Microorganisms are hypothesised to be passively dispersed by a number of mechanisms. Air currents can transport large numbers of microorganisms ($10^{18}$ – $10^{20}$) over vast distances (Griffin *et al.*, 2002) and are greatly dependent on wind velocity, however, it is normally only surface soils and sediments which can be dispersed. Ocean currents (for example, the Gulf Stream) are able to disperse microbes over short or vast distances, however, microorganisms can encounter extreme habitats *en route* to their new suitable location (Staley and Gosink, 1999). Some bacteria, such as members of the genus *Bacillus*, can form hardy life stages (e.g. spores) that are highly resistant to environmental stresses such as desiccation and better adapted to dispersal in the atmosphere; such stages can allow these organisms to disperse widely (Martiny *et al.*, 2006). Rain-fall also contributes to the passive dispersal, with surface soils and sediments washed away to new habitats. One obvious limiting factor to passive dispersal is geographical boundaries. Much fewer boundaries are present within the ocean than on land (i.e. mountains), therefore in principal there are fewer restrictions on the passive dispersal of microbes within the ocean, although this has never been unequivocally demonstrated and differences

in temperature and/or salinity as well as nutrient availability may in fact present substantial barriers in the ocean. As discussed previously, microorganisms can also be passively dispersed by vectors, in particular humans or birds, allowing rapid dissemination across the globe.

If free-living bacteria are cosmopolitan (i.e. they are found in more than one geographical location on Earth), this implies frequent migration from one location to another by one or more of the means described above; wind or water currents, or eukaryotic vectors. Therefore, though a combination of natural selection and rapid dispersal, two similar habitats on opposite sides of the globe can harbour the same bacterial species (Staley and Gosink, 1999). However, some studies have suggested that this is not always the case, and that spatial diversification is influenced more by geographic location than by the environment conditions. Whitaker *et al.*, (2003) investigated the biogeographic pattern of *Sulfolobus* sp. across numerous regions of Europe and North America. Rather than isolates clustering phylogenetically by environmental factors (i.e. *"...the environment selects."*) they instead cluster by geographic locale. In addition, region-specific cyanobacterial lineages have been observed from environments with a wide range of biochemical habitats (Papke and Ward, 2004). Therefore, geographic isolation can be of great importance for the development of global population structure, i.e. historic events shape the population (Martiny *et al.*, 2006).

With recent advances in molecular biology, it is now possible to examine the distribution of microbial species in natural communities, and to identify cosmopolitan "species". An extreme example of the spread of the widespread dissemination of a single genotype was reported by Breeder *et al.*, who noted that a microbial isolate from the North Sea crude oil fields was indistinguishable from an *Archaeoglobus fulgidus* strain isolated from an Italian hydrothermal system on the basis of DNA-DNA reassociation (Beeder *et al.*, 1994). However, this method of strain characterisation is relatively crude, and it is now possible to generate more definitive data on the basis of the nucleotide sequencing of multiple protein-coding genes. As mentioned previously (Chapter 1), MLSA is a rapid and robust method which allows the classification and identification of prokaryotes. The approach was used in the study by Whitaker *et al.*, in which nine chromosomal loci were sequenced

to assess the biogeographic pattern of the archaeon *Sulfolobus*. All strains were at least 99.8% identical in 16S rRNA, incriminating all strains as belonging to the same "species" (Whitaker *et al.*, 2003). However, across all nine loci the strains differ by up to 1.05%, indicating that there is a small, but significant level of genetic differentiation between populations which is related proportionally with geographic distance (Whitaker *et al.*, 2003). This highlights the increased power of MLSA in identifying fine-scale differences between populations.

### 6.1.5 *Vibrio* spp. global distribution

The study of marine bacteria, in particular the *Vibrio* genus, can address fundamental questions concerning patterns of microbial migration and gene flow. If local adaptation is minimal, such that it is not possible to predict the geographical origin of an isolate from its genotype, then the implication of high rates of global migration would heighten the risk to human health caused by *Vibrio* species. Transportation of *V. cholerae* strains via shipping is known to increase the risk of outbreaks; an epidemic in Peru in 1991 was traced to the importation of *V. cholarae* 01 biotype *El Tor* in a ships bilge water from the Far East (McCarthy and Khambaty, 1994). High rates of migration will consequently promote gene flow by recombination, because different species will have more chance of encountering each other; this may in turn lead to the emergence of novel hyper-virulent strains. A cholera pandemic throughout Thailand and the Indian sub-continent in 1993 was the result of the emergence of *V. cholerae* 0139, in which a section of the O-antigen gene cluster was replaced by recombination (Li *et al.*, 2002). Conversely, local adaptation can restrict gene flow because "intermediate" genotypes are less likely to be selectively tolerated. As a result, low levels of migration and local ecological adaptation will therefore reduce the risk of novel recombinants emerging in the environment.

The report by Thompson *et al.*, (2005) demonstrated an unprecedented degree of diversity within the *Vibrio* population sampled from the water column. The population examined represented a small sample of the diversity present within the *Vibrio* genus, with all isolates sharing >99% identity at 16S rRNA to *V. splendidus*. Thompson *et al.* reported extensive neutral sequence diversity within the *hsp*60 (heat shock protein) gene, as well as impressive heterogeneity of PFGE patterns and

genome size. They suggest that this diversity may reflect the heterogeneity of the water column, which would reduce the competition between clones. Predation may also play an important role by preventing highly adapted clones in dominating the population. Thompson *et al.* also suggest that associations with eukaryotic hosts are likely to be transient "random encounters", and that there is little evidence for host-specificity. However, as the isolates were drawn directly from the water column, the role of hosts in providing niches (and in turn driving diversity) is still poorly understood. However, the data available suggests that high degrees of variation are also observed in isolates recovered from single host species – a good example being the study of Buchrieser *et al.* who examined the diversity within *V. vulnificus* strains recovered from Oysters (Buchrieser *et al.*, 1995).

## 6.2 METHODS

### 6.2.1 Sampling

Sampling was conducted at four sites in New Zealand by Professor Paul Rainey and Annabel Gunn. These isolates were compared with those isolates from South Wales as described in Chapter 5. Four sampling sites in New Zealand were used (Fig 2); Bethells Beach and O'Neills Beach (Fig 3) on the East coast, and Westwell Beach and Narrow Neck Beach (Fig 4) on the West coast (each of the sites on either coast are approximately 1000 km apart by sea, but only 33 km apart by land). The two pairs of sites on each coast are located approximately one kilometre apart. Sampling within NZ was conducted once at each location during the summer months (Table 1). All New Zealand isolates were prepared and DNA sequenced by Annabel Gunn, and placed on the OSCART virtual laboratory system along with the UK isolates (http://herbaria.plants.ox.ac.uk/oscart/).

| Date | Location |
|---|---|
| 22/02/05 | Bethells Beach |
| 22/02/05 | O'Neills Beach |
| 24/02/05 | Westwell Beach |
| 24/02/05 | Narrow Neck Beach |

**Table 1** Sample site information.

### 6.2.2 Isolation of samples

Isolates from UK and NZ were recovered as described previously in Chapter 5 Section 2.2 *Preparation and Storage of Cell and DNA stocks*, however the target host organism of the NZ sampling was *Actinia tenebrosa,* which is a sister species of *A. equina* found in UK. The number of sea anemones sampled varied between samples sites, but the percentage of anemones positive for *Vibrio* was high and similar to that observed in the UK (Table 2). 10 isolates were also recovered from sea water in the East and West coasts.


**Figure 2** East and West coast sampling sites, New Zealand.


**Figure 3** Location of O'Neils Beach and Bethells Beach on NZ East coast.


**Figure 4** Location of Westwell Beach and Narrow Neck Beach on NZ West coast.

| Sample site | Source | No. samples | % positive for *Vibrio* spp. |
|---|---|---|---|
| O'Neils | *A. tenebrosa* | 20 | 80 % |
| Beach | Sea water | 10 | 100 % |
| Bethells | *A. tenebrosa* | 20 | 90 % |
| Beach | Sea water | - | - |
| Westwell | *A. tenebrosa* | 14 | 85.7 % |
| Beach | Sea water | 10 | 100 % |
| Narrow | *A. tenebrosa* | 14 | 92.9 % |
| Neck Beach | Sea water | - | - |

**Table 2** Number of host organisms sampled at each sampling date.

6.2.3 Isolate nomenclature

Each NZ *Vibrio* isolate was assigned a reference number to allow easy the identification of samples. For reference:

- 1 – 50 Bethells Beach
- 51 – 99 O'Neils Beach
- 100 – 149 Narrow Neck Beach
- 150 – 197 Westwell Beach
- 198 – 207 O'Neils Beach (sea water samples)
- 208 – 217 Westwell Beach (sea water samples)

6.2.4 Choice of genes, primer design & PCR conditions

PCR and sequencing of three genes (*recA*, *ompK* and *mdh*) were carried out as described previously (Chapter 2 Section 2.4 *Methods for Polymerase Chain Reaction and Sequencing* & Chapter 5 Section 2.4 *Choice of genes, primer design & PCR conditions.*

6.2.5 Nucleotide sequence analysis

-Sequence editing and alignment

As described in Chapter 2 Section 2.5 *Nucleotide Sequence Analysis.*

.

6.2.6 Phylogenetic analysis

-*recA* species assignment

As described in Chapter 5 Section 2.6 *Phylogenetic analysis.*

-MEGA

As described in Chapter 2 Section 2.6 *Phylogenetic analysis.*

-Bayesian inference of phylogeny

As described in Chapter 2 Section 2.6 *Phylogenetic analysis.*

The following parameters were used to create the Bayes block:

- number of generations = 6,000,000
- number of chains = 4
- sampling frequency = 1000
- print frequency = 100
- burnin = 20%.

### 6.2.7 Tests for recombination

-Tests of neutrality

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Sawyer's Run Test

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Population-scaled recombination rate (ρ)

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-DnaSp

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

-Splits decomposition

As described in Chapter 2 Section 2.7 *Tests for Recombination.*

### 6.2.8 Other methods of analysis

-eBURST

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

-Datamonkey

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

-Classification Index

As described in Chapter 2 Section 2.8 *Other methods of analysis.*

## 6.3 RESULTS

### 6.3.1 Summary of the dataset

In order to directly compare the sample of isolates from New Zealand with those from the UK, the same three loci, *recA, mdh* and *ompK*, were sequenced in the 105 NZ isolates; combined with the 180 UK isolates this resulted in a total dataset of 285 isolates. The sequences generated in NZ were marginally shorter than those generated in the UK, and the UK sequences were trimmed accordingly. The final length of the alleles after the removal of gaps was 387 bp *(ompK)*, 390 bp *(mdh)* and 612 bp (*recA*).

### 6.3.2 Species differences between the UK & NZ sites

As for the previous chapter, species assignments were made on the basis of Blastn similarity (E) scores for each query *recA* sequence against all the *Vibrio recA* sequences in GenBank (see Chapter 5 Section 2.6 *Phylogenetic analysis*, and Appendix D for strain details). Fifteen *Vibrio* species were identified from the NZ isolates on the basis of *recA* sequences. Figure 5 and Table 3 shows the frequency of species observed within the UK and NZ sites. The isolates from NZ are more diverse, corresponding to twelve 12 named species, compared to 8 from the UK. Within each site a single dominant species contributes to the population (UK=*V. cyclitrophicus* (50% of isolates), NZ=*V. pomeroyi* (47.6% of isolates). Only 5 species of *Vibrio* (*V. harveyi, V. kanaloae, V. pomeroyi, V. tasmaniensis* & *V. diabolicus*) were observed within both the UK and NZ sites. The remaining *Vibrio* species were isolated exclusively from either the UK (*V. cyclitrophicus, V. splendidus, V. lentus*) or NZ (*V. fortis, V. chagasii, V. fortis/chagasii, V. campbellii, V. alginolyticus, P. damselae*). The differences between the two sites suggest limited migration of species between the UK and NZ, and that the populations are influenced by local adaptation.

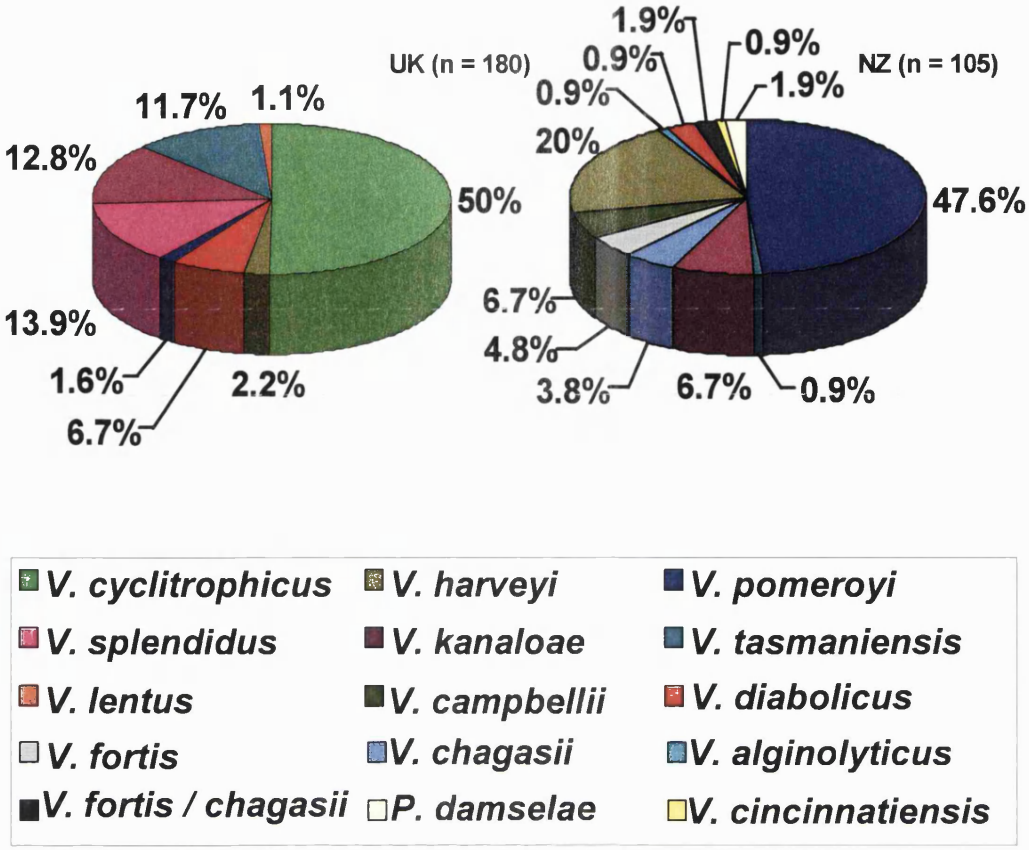



**Figure 5** Percentage diversity within each sample population.

| *recA* species assignment | Number of isolates | |
|---|---|---|
| | UK (%) | NZ (%) |
| *V. cyclitrophicus* | 90 (50) | - |
| *V. splendidus* | 25 (13.9) | - |
| *V. diabolicus* | 12 (6.7) | 3 (2.9) |
| *V. tasmaniensis* | 21 (11.67) | 1 (0.9) |
| *V. pomeroyi* | 3 (1.7) | 50 (47.6) |
| *V. harveyi* | 4 (2.2) | 21 (20) |
| *V. kanaloae* | 23 (12.8) | 8 (7.62) |
| *V. lentus* | 2 (1.1) | - |
| *V. chagasii* | - | 4 (3.81) |
| *V. campbellii* | - | 7 (6.7) |
| *V. fortis* | - | 5 (4.8) |
| *V. alginolyticus* | - | 1 (0.9) |
| *V. fortis/chagasii* | - | 2 (1.9) |
| *V. cincinnatiensis* | - | 1 (0.9) |
| *P. damselae* | - | 1 (0.9) |

**Table 3** Number of *Vibrio* species isolated from the UK and NZ sites.

6.3.3 Species differences between the East and West coasts of New Zealand

Marked differences were observed between the East and West coast sampling sites at North Island, NZ (Table 4). Of the twelve *Vibrio* species observed within NZ, only 5 were common to both the East and West coast (*V. pomeroyi, V. harveyi, V. chagasii, V. campbellii* & *V. fortis*). The majority of these species appear to be evenly distributed between the two coastal sites, with the exception of *V. pomeroyi,* in which a two-fold difference is observed at the East coast. Four *Vibrio* species were isolated exclusively from the East coast of NZ (*V. diabolicus, V. tasmaniensis, V. cincinnatiensis* & *V. kanaloae*) and three from the West coast of NZ (*V. alginolyticus, V. fortis/chagasii* & *P. damselae*).

| *recA* species assignment | Number of NZ isolates | |
|---|---|---|
| | East coast (%) | West coast (%) |
| *V. diabolicus* | 3 (2.9) | - |
| *V. tasmaniensis* | 1 (0.9) | - |
| *V. pomeroyi* | 34 (32.4) | 17 (16.2) |
| *V. harveyi* | 8 (7.6) | 12 (11.4) |
| *V. kanaloae* | 8 (7.6) | - |
| *V. chagasii* | 3 (2.9) | 1 (0.95) |
| *V. campbellii* | 2 (1.9) | 5 (4.8) |
| *V. fortis* | 3 (2.9) | 2 (1.9) |
| *V. alginolyticus* | - | 1 (0.9) |
| *V. fortis/chagasii* | - | 2 (1.9) |
| *V. cincinnatiensis* | 1 (0.9) | - |
| *P. damselae* | - | 2 (1.9) |

**Table 4** Number of *Vibrio* species isolated from East (n=63) and West coast (n=42) of North Island, NZ.

6.3.4 Sequence parameters

Various sequence parameters for the three genes within the UK population, NZ population and combined UK and NZ population are given in Table 5. The dS/dN ratio (per site synonymous base substitutions / per site nonsynonymous base substitutions) was calculated to assess the strength of purifying selection within each gene. The dS/dN ratios of the two housekeeping genes are high in both populations, reflecting a predominance of synonymous changes. As discussed in Chapter 5, for the UK data, *recA* is the most evolutionary conserved gene in that it exhibits both the

highest dS/dN ratio (33.1) and the lowest average pairwise divergence ($\pi$). Interestingly however, this gene also exhibits the highest number of different alleles of the three genes (59), although the high number of synonymous changes means that this allelic diversity is likely to have low functional relevance. *mdh* also exhibits a high dS/dN ratio (24.9) and a similar average pairwise diversity as *recA* which again is consistent with high levels of purifying selection.

When the data from New Zealand are considered, the dS/dN ratios for *recA* and *mdh* remain high, but interestingly in this case the ratio for *mdh* (34.6) is higher than for *recA* (22), thus the figures for these two genes are approximately inverted between the UK and NZ populations. Using a chi-sq test this difference is not statistically significant ($P > 0.05$), hence the importance of this observation is unclear. Clearly, however, the differences in the ratio between these two genes do not reflect differences in selection pressure as determined by gene function (as the ratio is reversed between the two sites). The differences may reflect sampling artefacts or possibly stochastic differences between the populations owing to recent recombinational replacements introducing a large amount of synonymous diversity into one or the other of these genes. When both UK and NZ populations are analysed together, the differences cancel out, and the *recA* and *mdh* exhibit very similar dS/dN ratios (28.54 and 29.47 respectively).

In contrast, the dS/dN ratio for *ompK* is consistently much lower in both the UK (4.1) and NZ (4.4) data, confirming that this gene exhibits the weakest levels of purifying selection of the three loci. In addition to the high proportion on non-synonymous changes, *ompK* is also the most variable of the three loci; when both populations are considered >50% of the nucleotide sites are variable, and the average pairwise divergence is 17.5%. This is consistent with the action of diversifying selection on this gene, which might be expected due to its role as a phage receptor. Rapid, evolution of this gene would confer frequency-dependent resistance to the predominant phage in the environment.

**UK population**

| Locus | Fragment size (bp) | No. alleles | % Variable sites | $d_S/d_N$ | π | θ | R | Tajima's D | Fu & Li F | Fu & Li D |
|---|---|---|---|---|---|---|---|---|---|---|
| *recA* | 612 | 59 | 28.99 | 33.13 | 0.065 | 39.65 | 1.90 | P > 0.01 | P > 0.01 | P > 0.01 |
| *mdh* | 390 | 49 | 44.62 | 24.91 | 0.066 | 25.60 | 0.001 | P > 0.01 | P > 0.01 | P > 0.01 |
| *ompK* | 387 | 56 | 56.33 | 4.12 | 0.150 | 58.09 | 3.70 | P > 0.01 | P > 0.01 | P > 0.01 |

**NZ population**

| Locus | Fragment size (bp) | No. alleles | % Variable sites | $d_S/d_N$ | π | θ | R | Tajima's D | Fu & Li F | Fu & Li D |
|---|---|---|---|---|---|---|---|---|---|---|
| *recA* | 612 | 57 | 35.67 | 22.05 | 0.099 | 60.64 | 8.90 | P > 0.01 | P > 0.01 | P > 0.01 |
| *mdh* | 390 | 54 | 35.13 | 34.59 | 0.124 | 15.70 | 15.7 | P > 0.01 | P > 0.01 | P < 0.02 |
| *ompK* | 387 | 53 | 56.07 | 4.37 | 0.176 | 68.21 | 29.3 | P > 0.01 | P > 0.01 | P > 0.01 |

**UK & NZ population**

| Locus | Fragment size (bp) | No. alleles | % Variable sites | $d_S/d_N$ | π | θ | R | Tajima's D | Fu & Li F | Fu & Li D | $F_{ST}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *recA* | 612 | 113 | 39.09 | 28.54 | 0.088 | 53.93 | 8.50 | P > 0.01 | P > 0.01 | P > 0.01 | 0.219 |
| *mdh* | 390 | 99 | 46.67 | 29.47 | 0.098 | 38.37 | 4.30 | P > 0.01 | P > 0.01 | P > 0.01 | 0.199 |
| *ompK* | 387 | 109 | 62.27 | 4.04 | 0.175 | 67.649 | 15.7 | P > 0.01 | P > 0.01 | P > 0.01 | 0.165 |

**Table 6** Nucleotide sequence variation of the UK and NZ *Vibrio* populations

When average pairwise diversity ($\pi$) is considered, the NZ sample is clearly more diverse than the UK sample in all three genes. This is consistent with the observation that a broader range of named species were detected in this population as assigned on the basis of *recA* sequences. Similarly, although the total number of alleles detected (for all three genes combined) is identical between the two samples (164), the UK sample consists of a slightly larger number of strains (n=180) than the NZ sample (n=105), thus the average number of different alleles per isolate is higher for the NZ sample (1.56) than for the UK sample (0.91).

The increased diversity in the NZ population is likely to be a reflection of the fact that this sample was drawn from two distinct geographical areas, whereas the UK population was drawn from essentially a single site. When the NZ data is split into the two main regions on the East and West coast of North Island, the average pairwise diversity in *recA* for isolates from these two regions is 0.087 and 0.112 respectively. Both these values are greater than the average pairwise diversity within the UK population, indicating a greater level of diversity within both NZ sites in comparison to the UK site. The *recA* dS/dN values for NZ East and West coast isolates are 20.43 and 20.78, respectively. Both these values are considerably less than the UK dS/dN (33.13) and the combined UK and NZ dS/dN (28.54) values, again indicating both sites in NZ have high levels of diversity.

Watterson's population mutation rate ($\theta$) was calculated for each locus within the two populations (Watterson, 1975). *mdh* has the lowest population mutation rate across the 3 loci in both the UK and NZ populations, with the lowest level observed within the NZ population. *recA* and *ompK* both have much higher $\theta$ values within the NZ population than that of the UK population. Hudson's R parameter, an estimate of the level of recombination within a population (Hudson, 1987), suggests that the NZ *Vibrio* population has a higher rate of recombination than that of the UK population, which is consistent with the higher level of diversity observed within this population. *ompK* exhibits the highest rate of recombination for both the UK and NZ loci. Again, this is consistent with its role as a phage receptor and suggests that recombination is an important mechanism by which variation may be generated at this locus. With the UK loci, *mdh* has the lowest rate of recombination at 0.001 with *recA* having the lowest rate within the NZ population.

To assess the level of neutrality within the two populations, Tajima's D, and Fu & Li's D* & F* tests were performed (Fu and Li, 1993). Within the UK population, these tests failed to provide significant evidence of selection all three loci, ($P > 0.01$). This consistent with the high dS/dN ratios observed in the two house-keeping genes *recA* and *mdh*, but is a more surprising result for *ompK* which exhibits diversifying selection. Within the NZ population, only Fu & Li's D* test with *mdh* provided evidence of selection ($P < 0.02$), again a surprising result given the evidence discussed above that this gene is under purifying (stabilising) selection. Although generally not significant Tajima's D, and Fu & Li's D* & F* tests all result in negative statistics (data not shown), indicating an excess of variants at low frequency, which suggests that these rare alleles are most likely deleterious.

To address the extent of gene flow and genetic differentiation between the UK and NZ populations, Wrights inbreeding coefficient ($F_{ST}$) was calculated (Wright, 1951). If the genotypes are randomly distributed between the two populations, then $F_{ST}$ would be zero. However, if the abundance of genotypes between sub-populations is non-random (i.e. they are partially isolated), genetic variation in each sub-population will be less then the total population and $F_{ST}$ will increase to a maximum of 1.0 (Bamshad *et al.*, 2004). The $F_{ST}$ values for the three loci are all greater than 0.15 indicating geographic structure between the UK and NZ populations and limited gene flow. Interestingly, *ompK* has the lowest value for $F_{ST}$ reflecting the lower levels of population differentiation and higher rates of gene flow between the UK and NZ populations and between the two NZ populations. This is likely to be related to the assumption that homologous replacement is more likely to confer a selective advantage at this gene than at either of the two housekeeping genes, thus gene flow from unrelated lineages is more likely to be observed.

When Wrights inbreeding coefficient is calculated for the East and West coast of NZ (Table 6), there are clear differences when compared with the UK and NZ $F_{ST}$ values. As expected, East coast versus West coast NZ values for all three genes are much lower than the UK and NZ indicating there is more gene flow within the NZ population than between the UK and NZ populations. An $F_{ST}$ value $< 0.1$ is

observed in *recA* and *ompK* for the two NZ samples, which is compatible with very low genetic differentiation, whereas the corresponding value for *mdh* (0.11) suggests moderate genetic differentiation.

| Loci | $F_{ST}$ value | |
|---|---|---|
| | UK & NZ | East & West NZ |
| *recA* | 0.21852 | 0.08571 |
| *mdh* | 0.19938 | 0.11418 |
| *ompK* | 0.16503 | 0.06151 |

**Table 6** Wrights inbreeding coefficient ($F_{ST}$), comparing UK (n=180) and NZ (n=105) to NZ East (n=63) and West coast (n=42) *Vibrio* isolates.

6.3.5 Phylogenetic analysis

Neighbour-Joining trees *recA*, *mdh* and *ompK* from all isolates from both the UK and NZ populations combined are shown in Figure 6, sections a, b & c respectively. The trees are colour coded according to species assignments based on *recA* Blastn scores, and also coded according to their source (UK or NZ).

-The *recA* tree

As discussed above, the abundance and diversity of species varied between the UK and NZ samples. Furthermore, inspection of the *recA* tree also reveals clear differences between the UK and NZ isolates even within the 5 species which are observed in both locations (Table 3). As the tree containing all isolates is difficult to interpret in detail, the five species common to both UK and NZ are now considered in turn by examining separate neighbour joining trees corresponding only to isolates assigned to the species concerned.

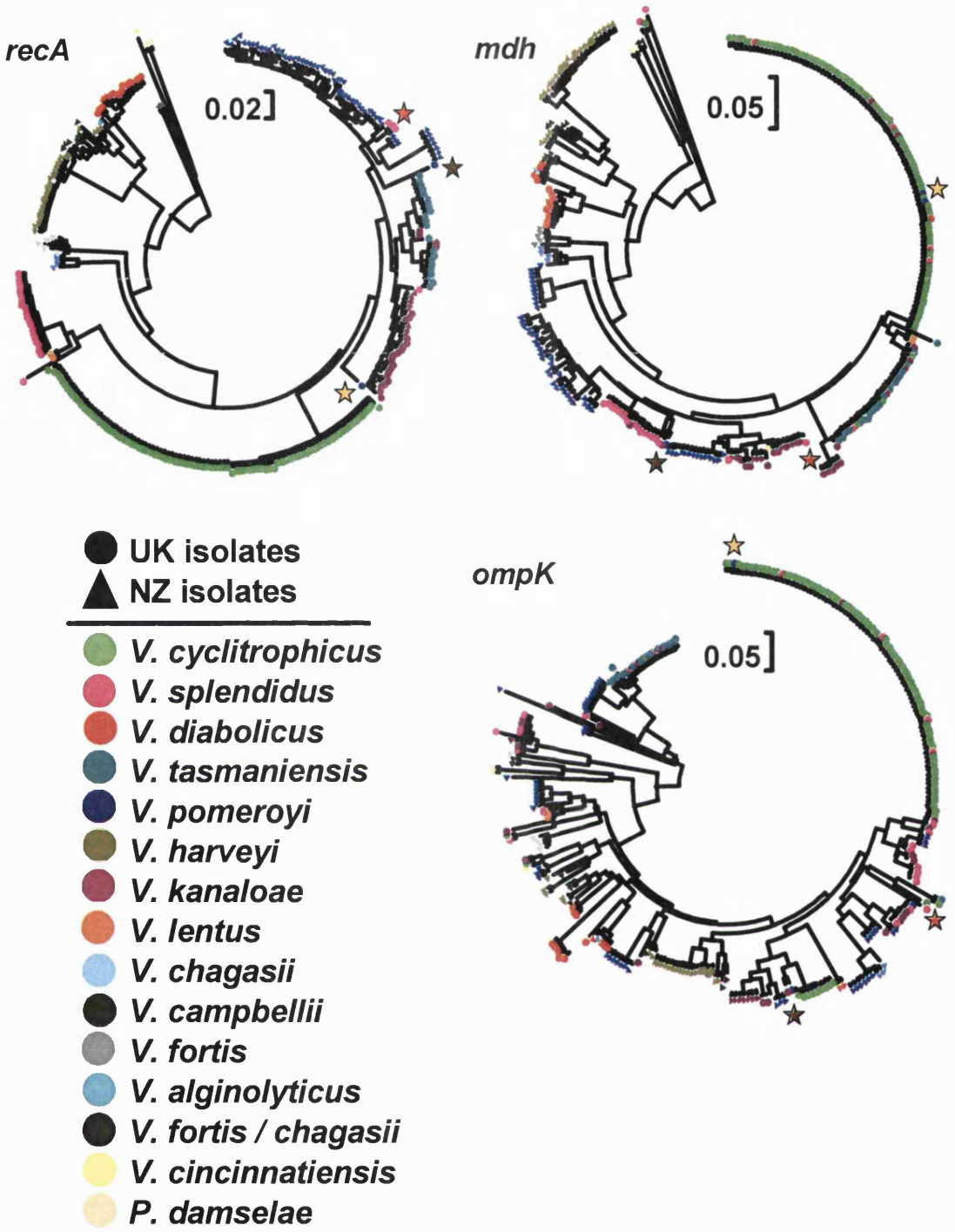



**Figure 6** Neighbour-joining trees for nucleotide sequence of UK (n=180) and NZ (n=105) MLSA loci for *Vibrio* spp. Coloured starts have been placed on each of the trees to allow easy identification of isolates discussed in further sections.

Firstly, the *V. harveyi* isolates are divided into two clusters, the larger of which contains isolates from both the UK (n=4) and from NZ (n=17), whereas the smaller cluster contains isolates from only NZ (n=3) (Fig. 7). This division is supported by a bootstrap score of 99 % indicating that this division is robust. The smaller cluster of exclusively NZ *V. harveyi* isolates is closely related to *V. campbellii* (bootstrap score of the node connecting the small *V. harveyi* cluster with *V. campbellii* is 99%).

A small but distinct cluster is also noted among the *V. kanaloae* isolates recovered from both continents, consisting of four isolates all recovered from the UK (bootstrap score 74 %; Fig. 8). All other *V. kanaloae* isolates belong to a single cluster containing 8 NZ isolates and 20 UK isolates (bootstrap score 81 %).

*V. tasmaniensis* was also noted in both the UK and NZ, although whereas 21 isolates were noted in the UK sample, only a single isolate from NZ was assigned as this species based on *recA* (Fig. 9). The *V. tasmaniensis* isolates are divided into two roughly equally sized clusters (n=12 & n=10) with the single NZ isolate corresponding to the larger of the two.

Little robust clustering is observed amongst the *V. diabolicus* isolates as indicated by the low bootstrap scores in figure 10. Possible exceptions are the clustering of 2 isolates from NZ (bootstrap score=99%) and 2 UK and one NZ isolate (89%).

*V. pomeroyi* is the final species observed in both the UK and NZ, but in this case 40 NZ isolates were assigned to this species and only 3 UK isolates (Fig. 6 & 11). Two distinct clusters are noted, the smaller one of which consists of 7 distinct (but identical) NZ isolates, separated from the other isolates of this species by a bootstrap score of 100%. One of the UK isolates corresponds with the larger *V. pomeroyi* clusters, showing an identical *recA* allele to NZ isolates, whereas the other 2 UK isolates are atypical (S2aA29vi & S2bA48v). Although they cluster together on the tree, this is likely to be an artefact as the bootstrap score is very low (32%) and each of these two isolates is on the end of a relatively long branch meaning that they are in fact fairly distinct from each other.

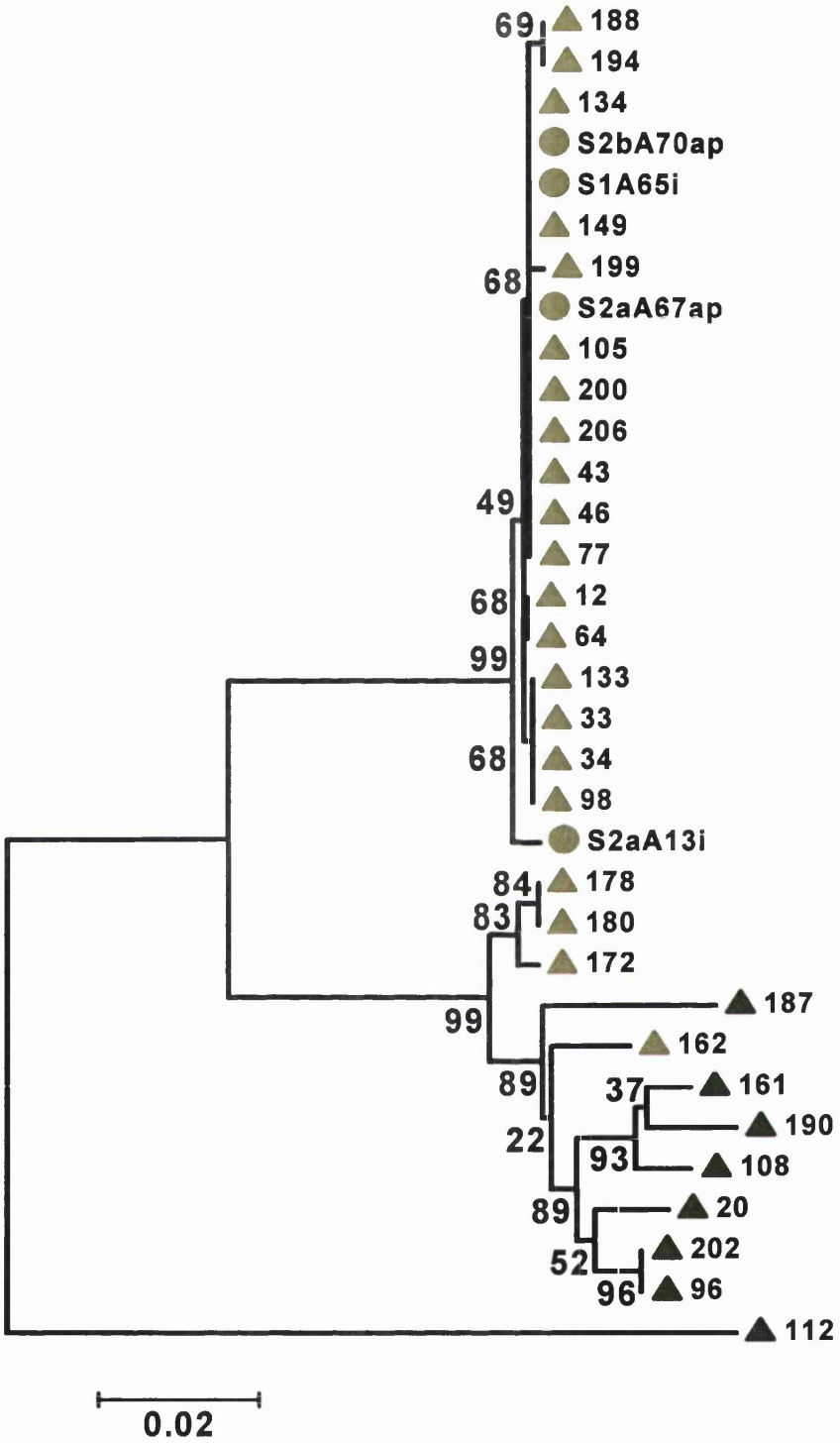



**Figure 7** Neighbour-joining tree showing *recA V. harveyi* (gold) and *V. campbellii* (green) isolates. Closed circles and closed triangles represent UK and NZ isolates respectively. Isolate 112, identified as *V. fortis/chagasii* was selected as an outgroup.

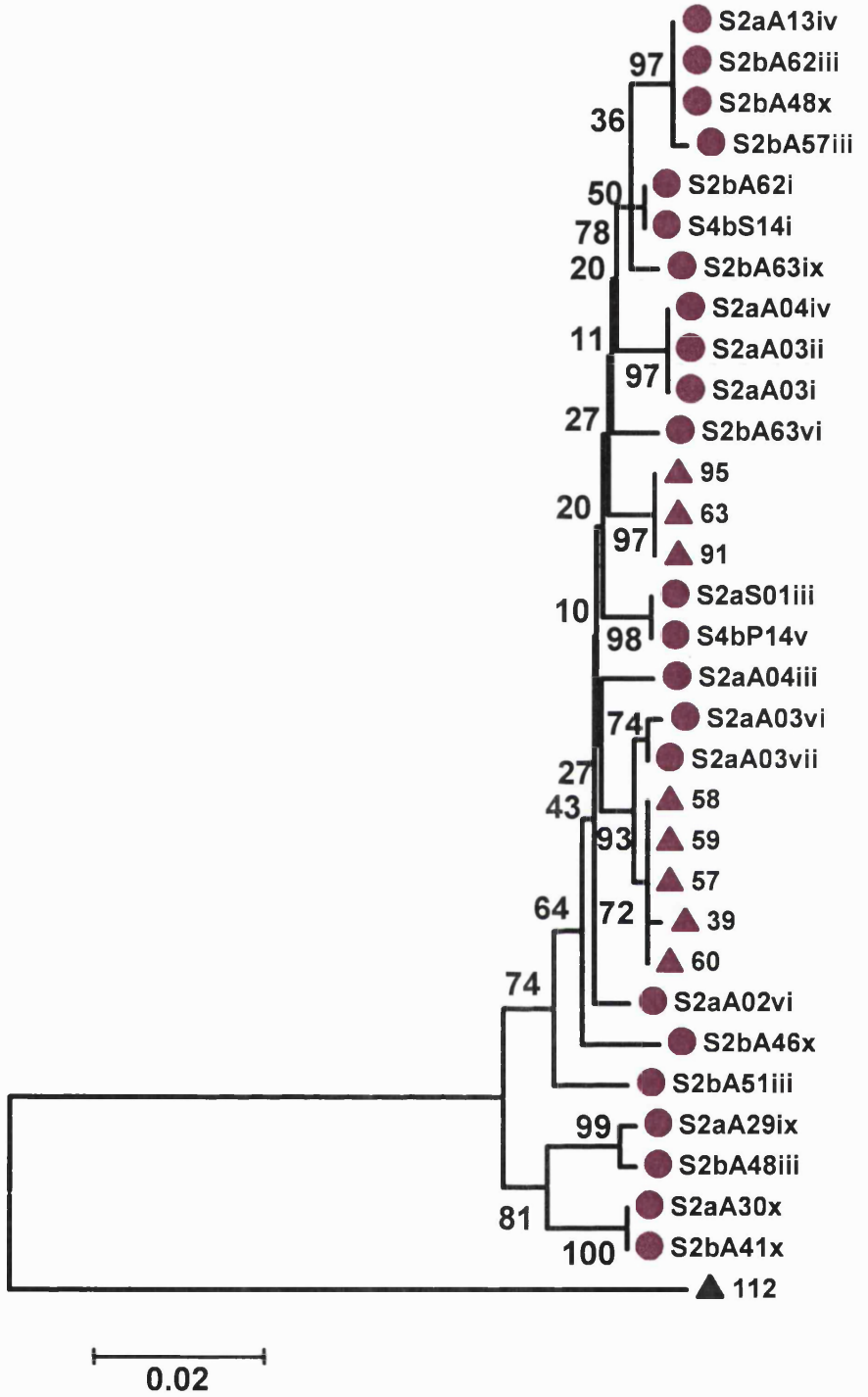



**Figure 8** Neighbour-joining tree showing *recA V. kanaloae* (purple) isolates. Closed circles and closed triangles represent UK and NZ isolates respectively. Isolate 112, identified as *V. fortis/chagasii* was selected as an outgroup.

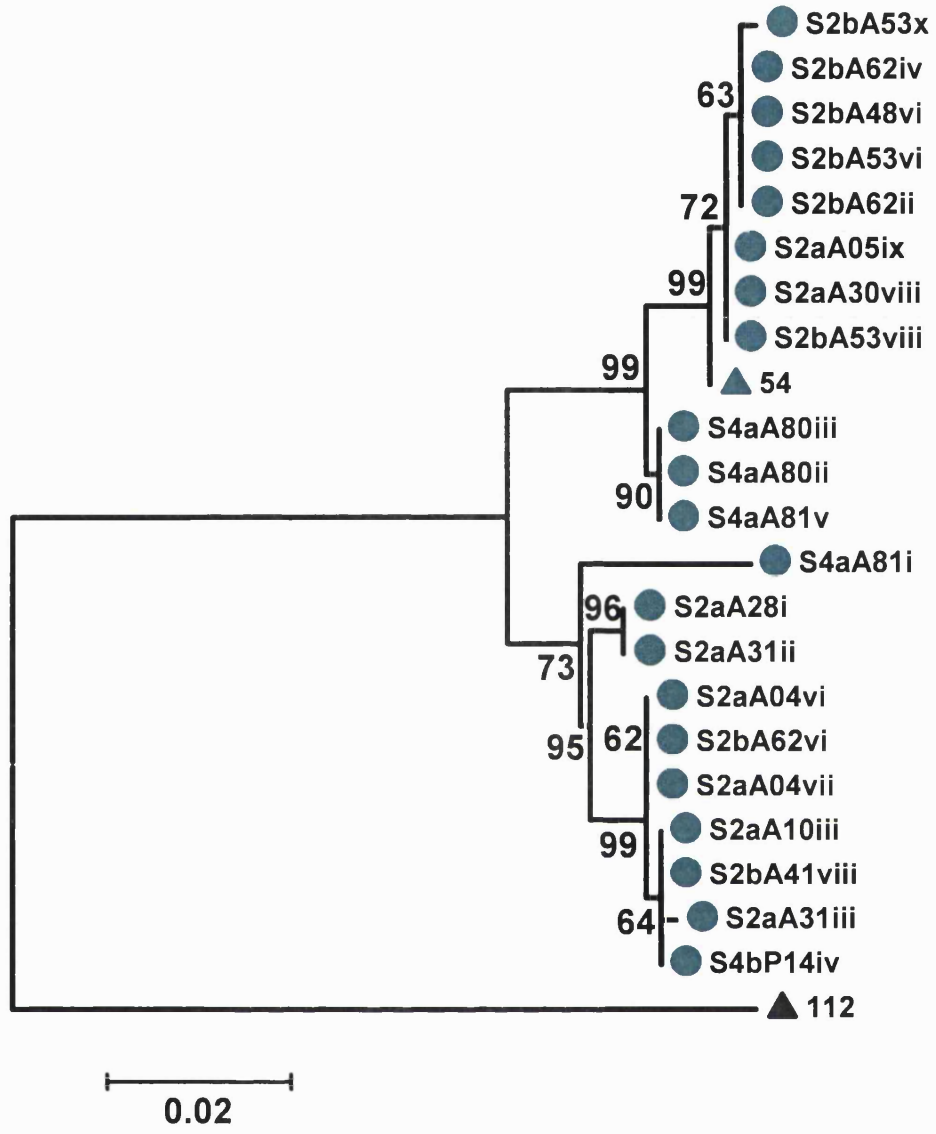



**Figure 9** Neighbour-joining tree showing *recA V. tasmaniensis* (turquoise) isolates. Closed circles and closed triangles represent UK and NZ isolates respectively. Isolate 112, identified as *V. fortis/chagasii* was selected as an outgroup.

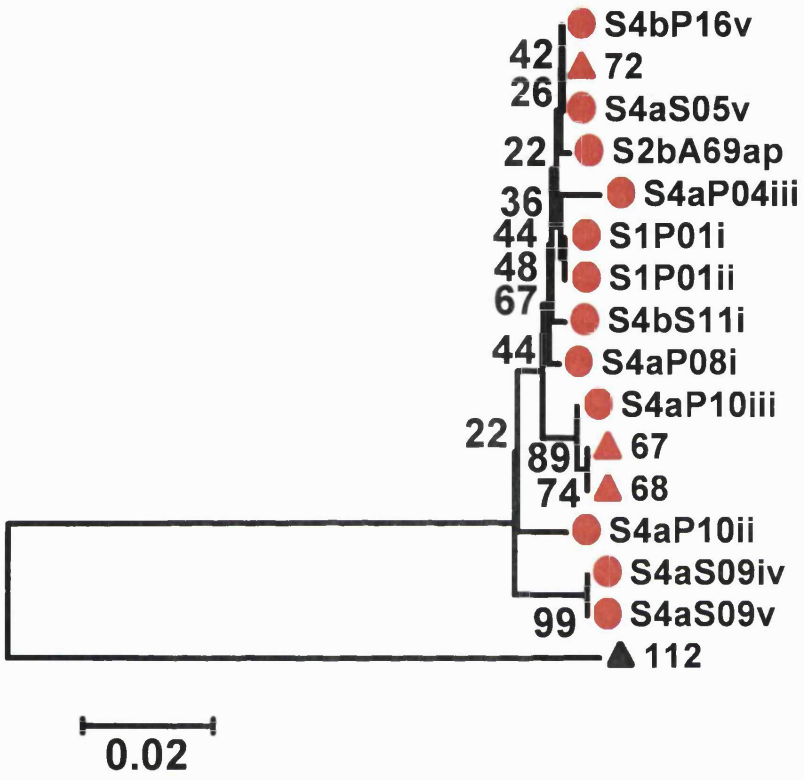



**Figure 10** Neighbour-joining tree showing *recA V. diabolicus* (red) isolates. Closed circles and closed triangles represent UK and NZ isolates respectively. Isolate 112, identified as *V. fortis/chagasii* was selected as an outgroup.

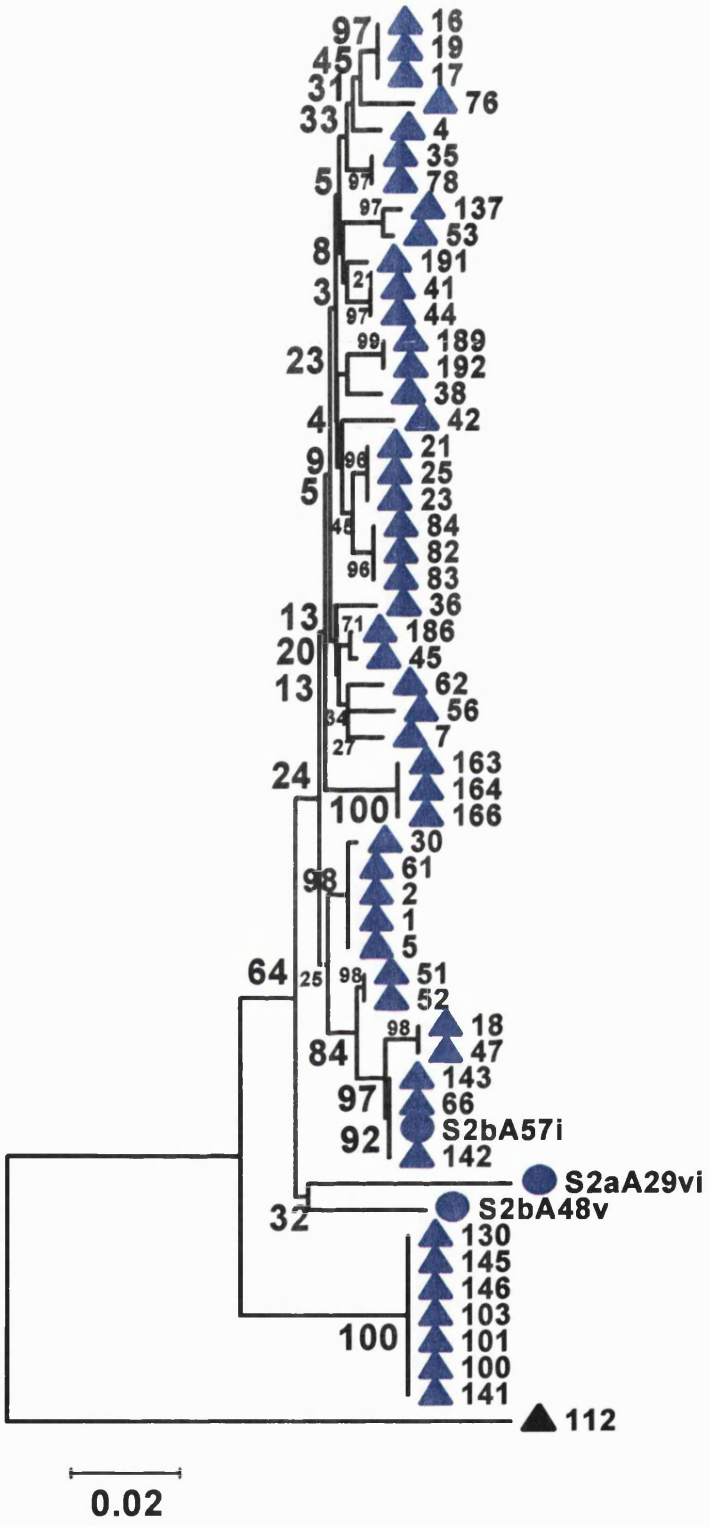



**Figure 11** Neighbour-joining tree showing *recA V. pomeroyi* (blue) isolates. Closed circles and closed triangles represent UK and NZ isolates respectively. Isolate 112, identified as *V. fortis/chagasii* was selected as an outgroup.

Although 5 species were noted in both the UK and NZ populations, all but *V. kanaloae* showed a pronounced predominance of strains in one or other of the sites. However, for most of the other species, there is still circumstantial evidence for the presence of localised clusters.

-Anomalous species assignments at *recA*

*V. splendidus* was only found within the UK population and forms two clusters on the tree, the first being a large cluster and the second a smaller group of only three isolates found within *V. pomeroyi*. Although this cluster is distinct it is clearly related to *V. pomeroyi* rather than *V. splendidus*, and the assignment of these strains as the latter species (highlighted in Figure 6 as a red stars) is most likely an artefact of relying on BLAST similarity searches rather than phylogenetic reconstruction for species assignment. A second irregularity is the single strain (S2bA48v) which has been assigned as *V. pomeroyi* on the basis of *recA*, but which is distinct from the main cluster of this species in the *recA* tree (highlighted in Figure 6 by yellow stars). Although clearly distinct on the basis of the *recA* sequence, this strain is identical to strains assigned as *V. cyclitrophicus* at both *mdh* and *ompK* and therefore appears to be a *V. cyclitrophicus* isolate which has acquired a divergent *V. pomeroyi*–like allele at *recA* by inter-species recombination. This is a clear example where species assignment made on the basis of *recA* would be inaccurate. A second such example is also highlighted in Figure 6 by brown stars. In this case strain S2aA29vi has been assigned as atypical *V. pomeroyi* on the basis of the *recA* sequence, but is identical to strains assigned as *V. kanaloae* at both *mdh* and *ompK*.

-The *mdh* tree

The *mdh* phylogeny provides further evidence concerning the reliability of species assignments made on the basis of *recA*. From a brief inspection of the *mdh* tree there appears to be broad agreement in terms of species assignment between the *recA* and the *mdh* genes, suggesting that, in general, *recA* is a reasonably reliable species marker. However, there are some interesting exceptions, and some notable inconsistencies between the topologies of the two trees.

Most notably, there are a number of strains which cluster with the large UK *V. cyclitrophicus* clade at *mdh*, but which have been assigned different species at *recA*

(*V. splendidus* n=3, *V. kanaloae* n=3, *V. pomeroyi* n=2, *V. diabolicus* n=1). One of these isolates is strain S2bA48v which, as discussed above, is assigned as *V. pomeroyi* on the basis of *recA*, but *V. cyclitrophicus* on the basis of *mdh* and *ompK*. Of the other nine isolates, 4 also cluster with *V. cyclitrophicus* on the basis of the *ompK* sequence. These are strains S2bA44viii, S2aA30ix and S2bA51iii (assigned as *V. splendidus* at *recA*), and strain S2bA62iii (assigned as *V. kanaloae* at *recA*). Therefore there are a total of 5 *V. cyclitrophicus* isolates which have been misassigned as other species on the basis of their *recA* sequences through the acquisition of a divergent allele via horizontal transfer. This is equivalent to approximately 5% of the *V. cyclitrophicus* isolates in the UK sample.

Similarly, there are a small number of isolates which have been assigned *as V. cyclitrophicus* on the basis of *recA*, but do not cluster with the main *V. cyclitrophicus* group at *mdh*. The UK *V. tasmaniensis* clade is more uniform at *mdh* than at *recA*, but also includes three isolates assigned as *V. cyclitrophicus* on the basis of *recA*. These strains (S2aA05vi, S2aA30viii & S2aA05x) also fall within the *V. tasmaniensis* clade on the *ompK* tree. A fourth isolate assigned as *V. cyclitrophicus* by *recA* is distinct at *mdh* and shown at the end of a long branch in approximately the 11 o' clock position on the *mdh* tree. The inconsistent positioning of these isolates is suggestive of a recombination event either at *recA* or at *mdh*. Evidence as to which of these two is most likely is provided by the *ompK* tree. In agreement with the *mdh* tree, none of these four isolates cluster with either of the two main *V. cyclitrophicus* groups at *ompK*. This means the inconsistency between *recA* and *mdh* is most likely to be the result of recombination at *recA*.

These data therefore provide a means by which the reliability of *recA* in species assignment can be quantified. A study into the abundance of *V. cyclitrophicus* within a population using *recA* as a sole marker would miss ~5% of the *V. cyclitrophicus* strains (i.e. they would be assigned as other species – "false negatives"). Similarly ~5% of the strains assigned as *V. cyclitrophicus* will be "false positives" and in fact would have been assigned as other species on the basis of other loci. It is noteworthy that comparisons with the *ompK* tree suggest that all nine discrepancies relating to the *V. cyclitrophicus* clade between *recA* and *mdh* are most likely to be due to recombination at *recA* rather than at *mdh*. This suggests that the

level of whole-gene inter-species recombination might in fact be unusually high at *recA*.

In addition to discrepancies in species assignment, there are marked topological differences between the *recA* and *mdh* tree. At *recA*, all the *V. pomeroyi* isolates group into a two closely related (variable) clusters. All but three of these isolates were isolated from NZ, and none of the three UK isolates which were assigned as *V. pomeroyi* by *recA* remained clustered with the NZ *V. pomeroyi* isolates at *mdh* or *ompK*; instead these strains clustered with *V. cyclitrophicus* (n=2) and *V. splendidus* (n=1) at *mdh*. This suggests that these isolates were incorrectly assigned as *V. pomeroyi* due to the acquisition of a *V. pomeroyi*-like allele at *recA*. A particular notable case is strain S2aA29vi (marked with a brown star on Figure 6), which possesses an atypical *recA* appearing to cluster with a distinct group of NZ *V. pomeroyi*. Analysis of the sequence confirms the distinctness of this allele, as it shows 51 nucleotide differences from this NZ clade. This is a further example of allelic replacement at *recA*, in this case a UK isolate has acquired a distinct allele which is most closely related to alleles observed in the NZ population. Thus, despite the clear genetic differentiation between the NZ and UK populations, there does appear to be some gene flow.

In contrast to *recA*, the *mdh* data divides the NZ *V. pomeroyi* strains into three distinct groups. Furthermore, the *V. splendidus* clade is interspersed between two of these *V. pomeroyi* groups at *mdh* but is clearly distinct at *recA*. This raises a further problem with using *recA* as a sole phylogenetic marker, namely that this gene lacks the resolution to identify biologically meaningful sub-groups within the named species. The *V. pomeroyi* group at *ompK* is even less coherent, as *V. pomeroyi* isolates are dispersed widely throughout the tree. In contrast, the conserved *V. cyclitrophicus* group remains reasonably coherent for all three genes; thus some named species are clearly phylogenetically "tighter" groups than others, possibly reflecting the different ages of the named species. The *V. cyclitrophicus* isolates in particular appear to have resulted from a recent clonal expansion, but even in this case the clade is split into two on the basis of the rapidly evolving *ompK*, and the two clades are sufficiently different that other species are interspersed between them.

In order to examine intra-species clustering within *mdh* in more detail, a separate Neighbour-Joining tree was constructed for *V. harveyi* isolates as for *recA*. Firstly, similar to *recA*, *V. harveyi* is split into two groups by *mdh* (Fig. 12). The larger of these is made is of UK (n=4) and NZ (n=15) isolates, whilst the smaller (again closely related to the *V. campbellii*) consists of the same four NZ isolates (172, 180, 162, 178) as seen for *recA* (Fig. 7). This consistency suggests that these four UK strains do in fact constitute a biologically meaningful subdivision, assigned as *V. harveyi* by *recA* BLAST scores, but falling at an intermediate position between *V. harveyi* and *V. campbellii* on the *recA* and *mdh* trees. For *recA*, one of these strains (162) falls within the *V. campbellii* cluster.

*V. kanaloae* is again divided into two clusters by *mdh*, where the larger cluster contains both UK and NZ isolates, as well as isolates assigned as *V. splendidus* (n=2) and *V. cincinnatiensis* (n=1) by *recA*; this is further evidence of indicative of interspecies recombination. In the case of the single *V. cincinnatiensis* isolate (shown in yellow and which falls in the middle of the larger *V. kanaloae* cluster in Figure 6), this strain appears more distinct from the *V. kanaloae* and *V. splendidus* strains at *recA* than at *ompK*, which again suggests that the allele at *recA* is aberrant, although the lack of structure in the *ompK* tree means that this is a less clear-cut case.

The *mdh* data split the *V. diabolicus* into two related clusters. Within the smaller cluster is a single isolate which was assigned as *V. alginolyticus* by *recA*, however, this isolate clusters with the *V. diabolicus* isolates at both *mdh* and *ompK* (Fig. 6). This again is further evidence of allelic replacement at *recA*.

Finally, alongside this grouping are three UK isolates, assigned as *V. splendidus* (n=2) and *V. cyclitrophicus* (n=1) on the basis of *recA*, showing marked diversity from the rest of the *mdh* tree and evidence of interspecies recombination events, potentially with *Vibrio* spp. which have not been sampled within this study.

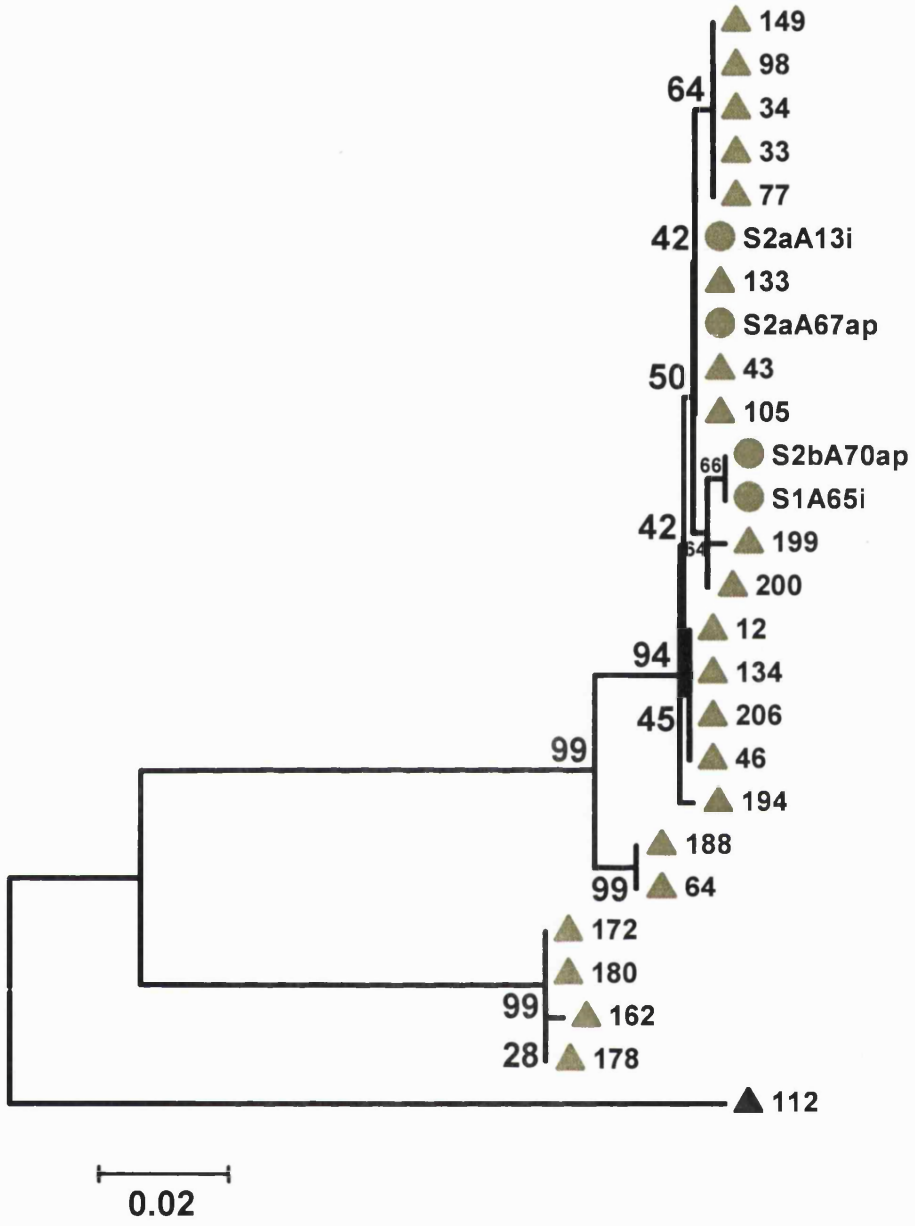



**Figure 12** Neighbour-joining tree showing *mdh V. harveyi* (gold) isolates. Closed circles and closed triangles represent UK and NZ isolates respectively. Isolate 112, identified as *V. fortis/chagasii* was selected as an outgroup.

-The *ompK* tree

*ompK* is by far the most divergent of the loci showing far less congruence to *recA* and *mdh* than these genes do to each other. Although most of the clusters do not contain multiple assigned species on the *ompK* tree, an exception is *V. splendidus* and *V. kanaloae* which are well distinguished by the housekeeping genes (particularly *recA*), but correspond to three clusters at *ompK*, each of which contains multiple isolates of each assigned species. Most notably, however, there are dramatic topological differences in the *ompK* tree compared to the trees from the housekeeping genes, in that individual species are less likely to correspond to single clusters, but instead form multiple clusters dispersed throughout the tree. Of particular note are the isolates corresponding to *V. pomeroyi*. On the *recA* tree, these isolates fall into two related clusters, for *mdh* they fall into 3 clusters (two of which are interspersed by the *V. splendidus* isolates), but at *ompK* the *V. pomeroyi* isolates fall into 7 major clusters (11 clusters including minor branches) which are no more closely related to each other than to isolates of other species.

There is also evidence of clustering dependent on geographical source on the *ompK* tree. For example, *V. kanaloae* is divided into two clusters at *mdh*, however with *ompK* this species is comprised of one main cluster containing the NZ isolates and then a number of smaller UK clusters closely associated with *V. splendidus*. *V. diabolicus* also exhibits clustering based on the country of isolation, with the main *V. diabolicus* cluster comprised of UK isolates and a smaller cluster of NZ isolates placed in a different locality on the tree.

More strikingly, however, the clusters of isolates corresponding to *V. pomeroyi* reflect isolation from the two sites in New Zealand. For example, considering all 11 *V. pomeroyi* clusters on the *ompK* tree, 6 exclusively contain isolates originating from the West coast, 3 exclusively from the East Coast and only 2 clusters contain both East and West coast isolates. This level of separation between the two NZ sites is not observed within any of the other loci.

### 6.3.6 Clonal diversification

Allelic profiles of each UK and NZ isolate were assigned on the basis of the nucleotide sequence of each locus, and the structure of the population examined using eBURST (Fig. 13). The current version of eBURST (version 3) allows the comparisons of two populations and in figure 13 the NZ isolates are highlighted by a green halo. The founder clones of each clonal complex are coloured blue and all sub-founders are coloured yellow for both UK and NZ isolates. The UK and NZ populations contain 172 unique STs with 77% (n=133) of the STs represented by only a single isolate and 4% (n=7) represented by a minimum of four isolates. The 172 STs were divided by eBURST into one major clonal complexes, five minor clonal complexes (containing between 4 and 7 STs), twelve doublets and the remaining 101 STs as singletons (single isolates which are not SLVs of any other isolates in the sample). The largest and most common clone was ST3, which accounts for 38 isolates (13.3% of all isolates) and this is the predicted founder of the major clonal complex (bootstrap score 82%). This clonal complex contains 23 STs and 88 isolates, all of which were recovered from the UK. The second largest complex is composed of 7 STs and 11 isolates, 10 of which were recovered from NZ (the exception being ST59). The predicted founder of this complex is ST172, although this assignment is not very robust (bootstrap score 61%). All of the other minor clonal complexes consist entirely of UK isolates (e.g. nt-CCs 9, 68 and 104), with the exception of nt-CC135 (bootstrap score 45%) which exclusively contains isolates from NZ. The founder of this minor clonal complex cannot be assigned with confidence, for example both ST135 and ST112 define 3 SLVs and 1 DLV and hence produce similar bootstrap scores. Similarly, STs 172 and 62 are equally parsimonious founders for CC172 (which is named as such, rather than CC62 arbitrarily). However, the assignment of ST3 as the founder of the large clonal complex is much more robust, and is also consistent the position of this ST in splits decomposition analysis of the complex as discussed below. This complex corresponds to the large clade of *V. cyclitrophicus* from the UK sample discussed previously. Figure 14 illustrates the UK and NZ population based on two alleles in common, and details of the variant alleles observed can be found in Appendix D.

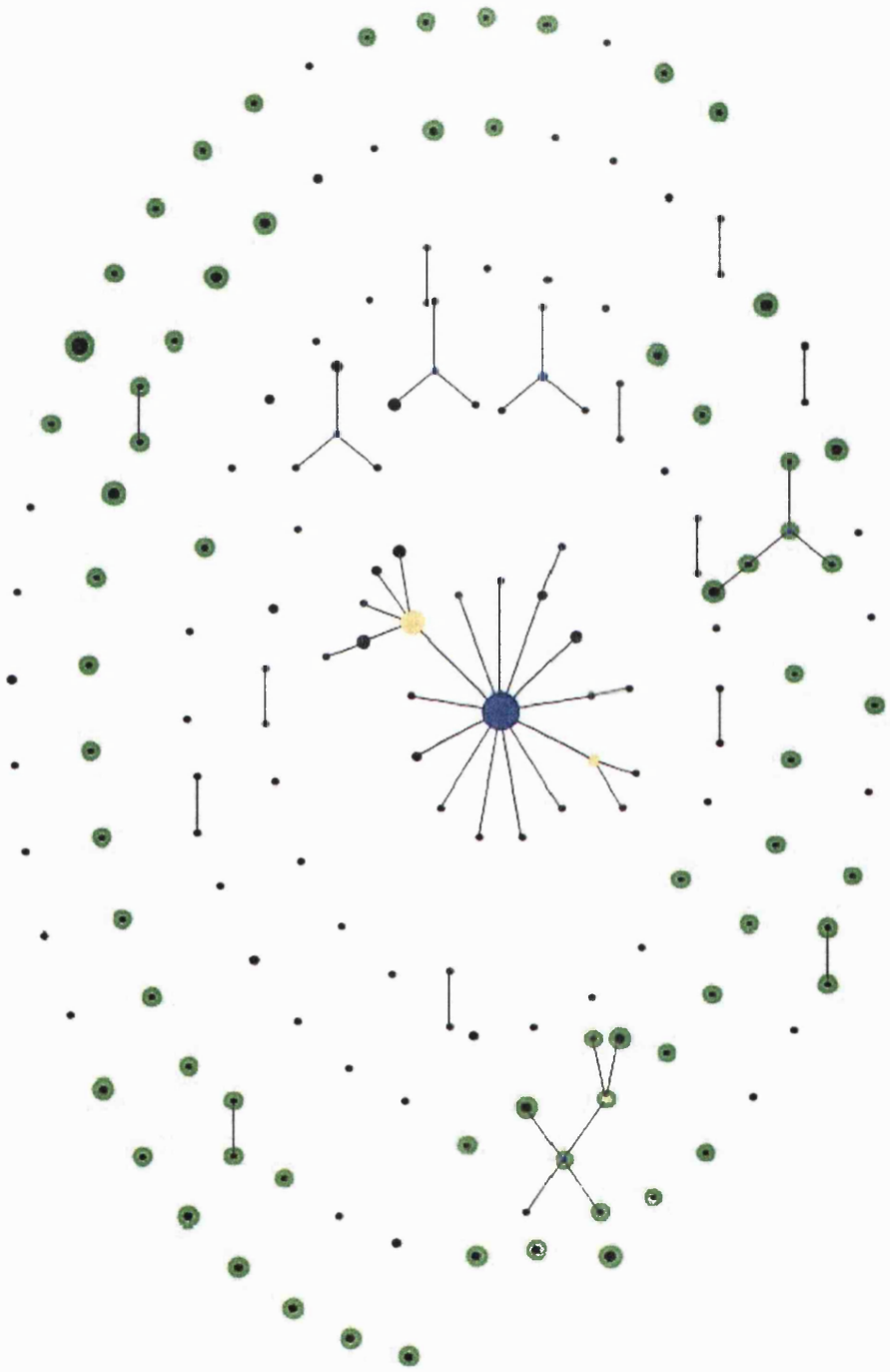


**Figure 13** A population "snapshot" showing the clusters of linked STs and unlinked STs for 180 UK and 105 NZ (green halo) *Vibrio* spp. using nucleotide sequence data (0/3 alleles in common). The predicted clonal ancestor is shown in blue, the subgroup founders shown in yellow. SLVs are shown as black lines. The sizes of the circles that represent each ST indicate their prevalence within the population.

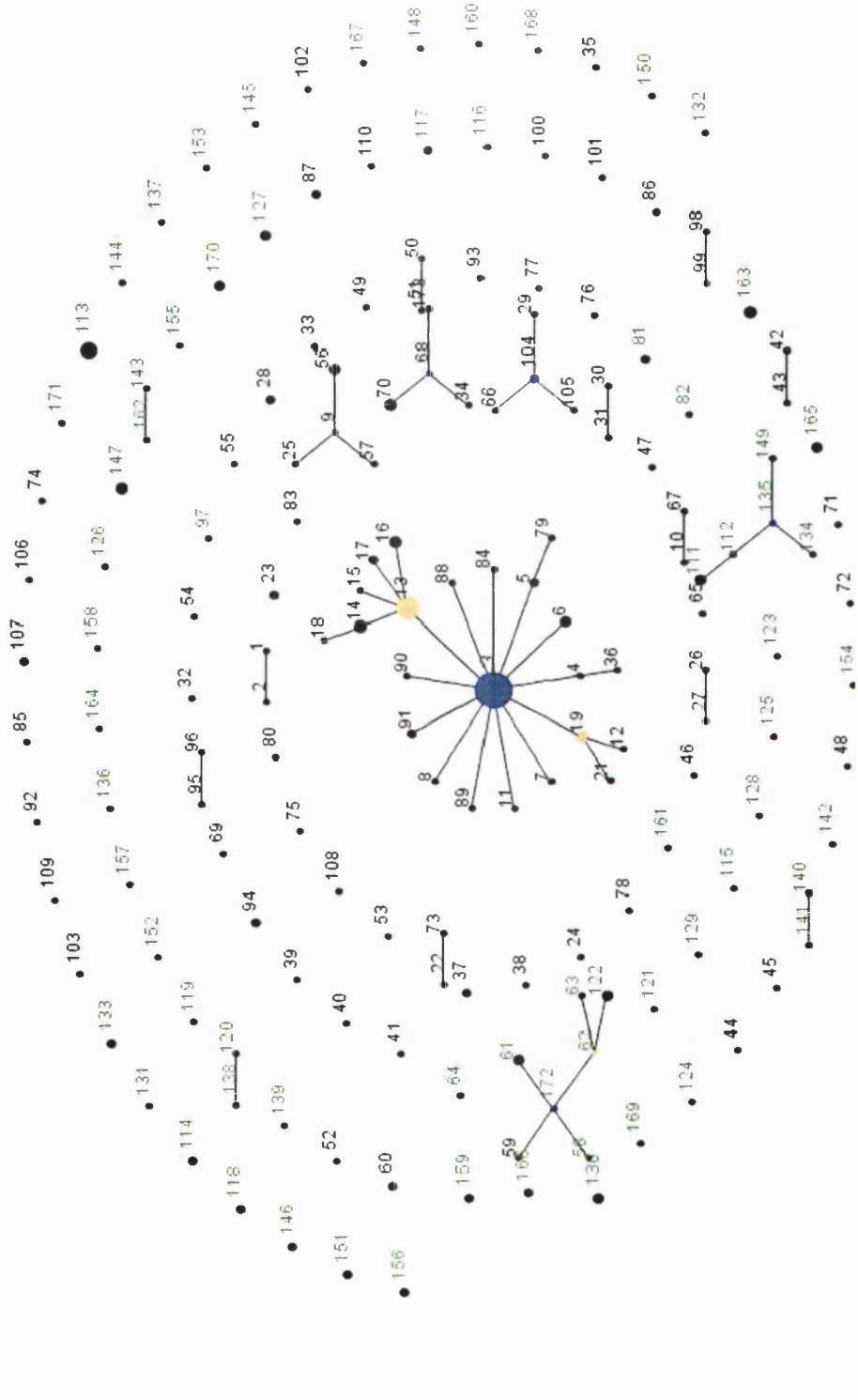


**Figure 14** Identification of 1 major clonal complex (na-CC3) and 5 minor clonal complexes (na-CC9, na-CC68, na-CC104, na-CC112 & nt-CC172). The size of the circles representing each ST indicates their prevalence within the population. eBURST figure constructed using 0/4 loci in common.

### 6.3.7 eBURST based on translated alleles

Multi-locus sequence typing and eBURST were originally designed for the analysis of single species, whereas the data presented here is representative of several distinct *Vibrio* species. The power of eBURST to make meaningful inferences concerning recent evolutionary pathways will be limited for data encompassing cross-genera comparisons because the proportion of isolate pairs corresponding to SLV links will be low. However, it is clear that much of the diversity present, at least in *recA* and *mdh,* is synonymous, and the use of translated alleles presents a means of filtering out much of the diversity to consider evolutionary pathways over longer time-scales.

The allelic profiles of each UK and NZ isolate were assigned on the basis of their amino acid sequence, and examined by eBURST (Fig. 15). A total of 110 STs were noted for both samples combined, with 63.6% (n=70) of all STs represented by only a single isolate, and 12.7% of all STs (n=14) represented by a minimum of four isolates. In order to cross reference the ST assignments based on nt- and aa- alleles, a table is presented in the Appendix D. Two major clonal complexes are noted using the translated alleles, along with four minor clonal complexes, one triplet, four doublets and eleven singletons. The NZ isolates have been highlighted by a green halo on Figure 15, which illustrates a clear distinction between the two populations. Not one clonal complex contains isolates from both the UK and NZ; each is composed exclusively of isolates from one or other of the two samples.

The largest clonal complex, aa-CC14, consists exclusively of UK isolates, and the founder of this complex is aa-ST14 (bootstrap support 72%). This ST corresponds to the nt-ST3 (the *V. cyclitrophicus* clade) with an additional 14 nt-STs also corresponding to aa-ST 3 (STs 2, 4, 13, 15, 17, 19, 21, 53, 75, 88, 91, 95, 98, 104). Notably, several of these nt-STs were unlinked to nt-CC3 on the basis of nucleotide alleles (53, 75, 95, 2, 104, 98), reflecting the fact that a large proportion of the diversity is synonymous. A major subgroup of aa-CC14 is aa-ST3 which corresponds to nt-STs 70, 101, 105, 68, 66, 56, 34, 29, 9, 7. None of these nt-STs are linked to nt-CC3, again reflecting the large proportion of synonymous change. aa-ST36 represents a third subgroup founder of this complex. The patterns of descent inferred by eBURST within aa-CC3 correspond closely to simple radial diversification, suggesting the use of translated alleles may be a powerful approach

in inferring patterns of descent from genus-wide MLSA data. A minor clonal complex founded by aa-ST53 is also noted, which, similar to aa-CC3, exists exclusively of UK isolates.

The NZ sample consists of a single major clonal complex, and 3 minor clonal complexes. The assigned founder of the major clonal complex is aa-ST67 (bootstrap support 63%). However, this assignment is not robust as this ST defines 13 SLVs, whereas 12 other STs within this complex define 12 SLVs. Furthermore, this assignment is not supported in terms of the frequency of the ST, which was only noted once, whereas aa-ST75 is noted in 12 isolates. Although it is not possible to deduce the founder with confidence in this case, forcing aa-ST75 as the founder makes little difference in the patterns within the group. This lack of resolution stems from the fact that only 3 loci were sequenced for these strains, and in fact 13/14 aa-STs belonging to this complex, which correspond to the isolates assigned as *V. pomeroyi*, are identical at *recA* and *mdh* (corresponding to alleles 13 and 1 respectively), and only differ at *ompK*. Thus this complex reveals a remarkable degree of non-synonymous divergence within *ompK* amongst a clone of isolates which show identical aa-alleles at the two housekeeping genes. The exception is aa-ST77 which exhibits *mdh* aa-allele 22, but similarly to aa-ST67 exhibits *ompK* allele 14. Hence when an alternative founder is assigned to this group (e.g. aa-ST-75, which is the most frequent aa-ST of this complex), aa-ST77 becomes a DLV of the founder with aa-ST67 as an intermediate SLV (Fig. 6a). Similarly, the founder of the minor NZ complexes aa-CC87 (bootstrap score 71%) is not supported by the frequencies of the STs. As aa-ST81 is only noted once, whereas aa-ST86 is noted in 12 isolates, it is possible that aa-ST86 may represent the true founder of this group (Fig. 16b). The predicted founders of the minor NZ complexes, aa-CC97 and aa-CC105, also lack robustness.

In summary, this analysis reveals significant differences between the NZ and UK samples. In the case of the UK population, the majority of the isolates cluster into a single well-defined group with a robust founder and sub-founder, whereas the NZ isolates belong to smaller distinct clusters where the founders cannot be assigned with such confidence, owing primarily to the rapid rate of allelic change at *ompK*. In contrast to the NZ complex aa-CC67, where all but one of the SLVs has changed

at *ompK*, of the 17 SLVs of the major UK clone aa-ST14, 5 have changed at *recA*, 3 have changed at *mdh*, and 8 have changed at *ompK*. The reasons for the relatively high rate of change at *recA* within the UK samples are unclear, but appear to be due to recombination rather than point mutation (phylogenetic analysis described above and SLV analysis below).

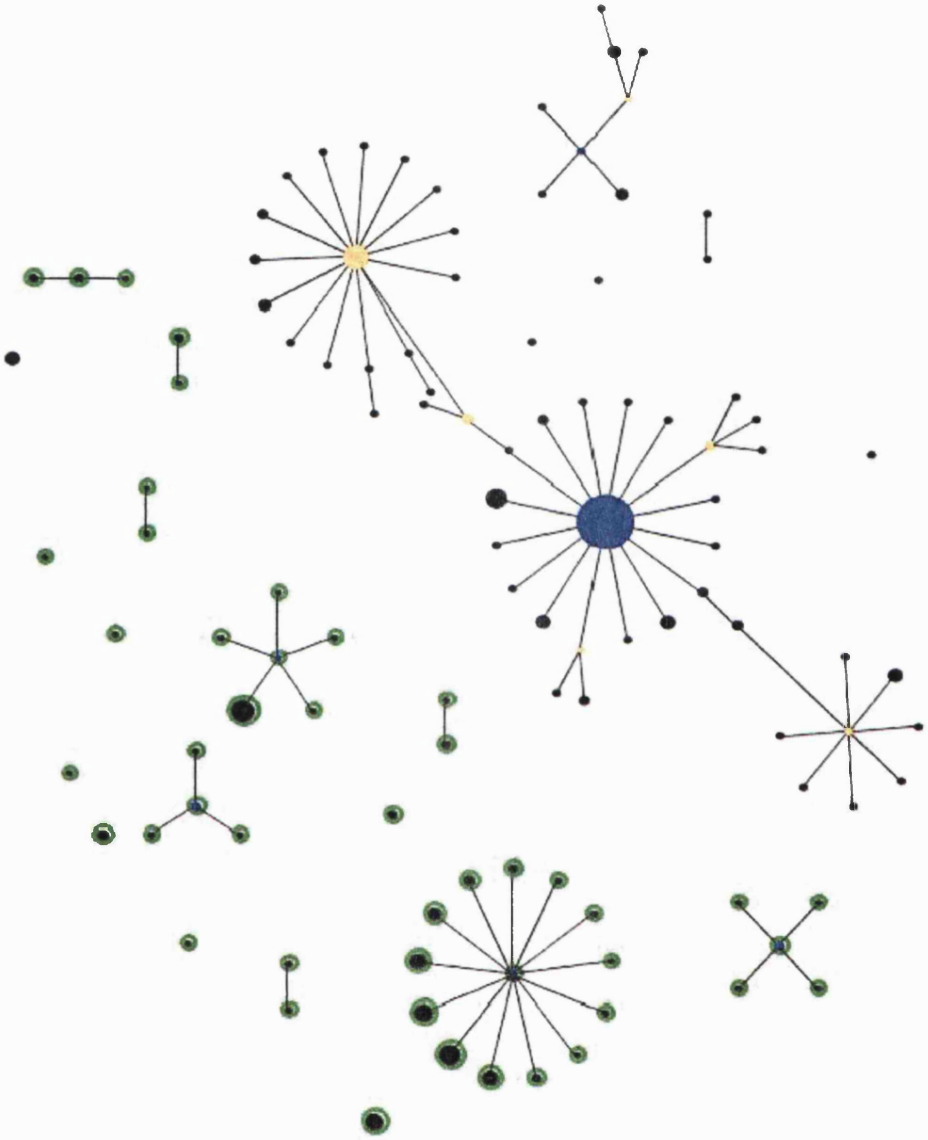


**Figure 15** eBURST applied to translated allelic profiles. A population "snapshot" showing the clusters of linked STs and unlinked STs for 180 UK and 105 NZ (green halo) *Vibrio* spp. using nucleotide sequence data (0/4 alleles in common). The predicted clonal ancestor is shown in blue, the subgroup founders shown in yellow. SLVs are shown as black lines. The sizes of the circles that represent each ST indicate their prevalence within the population.

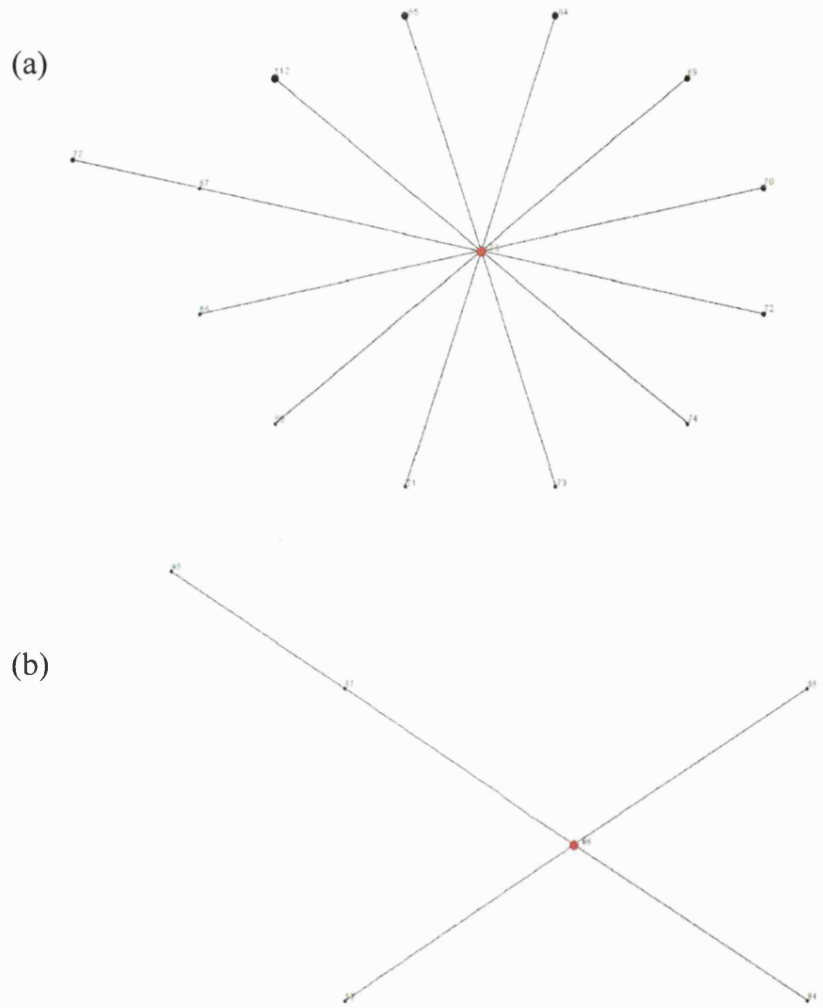



**Figure 16** Alternative founder STs for clonal complexes. Based on frequency, ST75 (a) and ST86 (b) should be the founder STs for each clonal complex.

### 6.3.8 Estimates of the R/M parameter

As mentioned in Chapter 5 it is possible to estimate whether SLVs and DLVs have arisen by recombination or by point mutation (Feil *et al.*, 2003). Within the UK and NZ samples, the majority of SLVs (37.1%) originate from *de novo* mutation (Table 7), i.e. a single, novel, nucleotide change in the particular SLV. However, in contrast, 45.7% of the nucleotide changes involved more than 12 nucleotide changes, with the greatest difference observed between ST135 and ST134 (66 bp). This strongly bimodal distribution indicates that at least half of the alleles have diversified by relatively diverse recombinational replacements, perhaps even from a different species.

This analysis also allows comparisons regarding the extent to which recombination has affected the three loci and reveals similar rates of mutation, but quite dramatic differences in the rates of recombination. The number of SLVs assigned as having arisen by point mutation are 2, 1 and 1 for *recA*, *mdh* and *ompK* respectively, hence the loci are exhibiting similar levels of purifying selection with regard to the purging of slightly deleterious mutations, at least over the very initial stages of clonal diversification. However, the corresponding values for SLVs putatively assigned as having arisen by recombination are 17, 3 and 10 for *recA*, *mdh* and *ompK* respectively. These figures correspond to R/M ratios of 8.5 (*recA*), 3 (*mdh*) and 10 (*ompK*) (Table 8). Surprisingly, therefore, the probability of an allele being affected by recombination rather than mutation is just as high in *recA* than in *ompK*, and very little recombination has been detected in SLVs at *mdh*. This is also consistent with the phylogenetic analysis which indicated that recombination has been more frequent at *recA* than at *mdh*, as in cases where the trees from these two genes are not consistent, the *ompK* tree generally supports the *mdh* topology. The higher rate of recombination at *recA* compared to *mdh* is difficult to explain in terms of gene function, as both are considered essential housekeeping genes, however, the high rate of synonymous substitution within *recA* means that recombination events are likely to be selectively neutral and do not affect this gene's overall function.

| No. of nucleotide differences | No. of SLVs | ST of SLVs |
|---|---|---|
| 1 bp | 13 | ST3→ST4, ST5, ST11, ST13, ST84<br>ST13→ST15, ST16, ST17, ST18<br>ST19→ST12, ST21<br>ST104→ST105<br>ST112→ST149 |
| 2 bp | 2 | ST3→ST19, ST88 |
| 3 bp | 2 | ST3→ST89, ST91 |
| 6 bp | 2 | ST3→ST90<br>ST104→ST66 |
| 12 bp | 1 | ST68→ST34 |
| 13 bp | 1 | ST112→ST135 |
| 18 bp | 1 | ST104→ST29 |
| 24 bp | 1 | ST68→ST70 |
| 25 bp | 2 | ST1→ST2<br>ST68→ST173 |
| 36 bp | 1 | ST5→ST79 |
| 46 bp | 1 | ST9→ST56 |
| 47 bp | 1 | ST9→ST57 |
| 51 bp | 1 | ST4→ST36 |
| 53 bp | 1 | ST22→ST73 |
| 54 bp | 1 | ST3→ST7 |
| 56 bp | 3 | ST3→ST6, ST8<br>ST13→ST14 |
| 66 bp | 1 | ST135→ST134 |

**Table 7** Number of nucleotide base pair differences throughout the SLVs.

| ST of clonal ancestor | ST of SLV | Variant locus in SLV | Ancestral allele | SLV allele | No. of bp differences | Mode of nucleotide change |
|---|---|---|---|---|---|---|
| 3 | 84 | *recA* | 2 | 37 | 1 | point mutation |
| 3 | 7 | *ompK* | 55 | 104 | 54 | recombination |
| 3 | 89 | *recA* | 2 | 42 | 3 | recombination |
| 3 | 90 | *recA* | 2 | 43 | 6 | recombination |
| 3 | 5 | *ompK* | 55 | 77 | 1 | recombination |
| 3 | 6 | *ompK* | 55 | 78 | 56 | recombination |
| 3 | 4 | *ompK* | 55 | 69 | 1 | recombination |
| 3 | 19 | *recA* | 2 | 4 | 2 | recombination |
| 3 | 8 | *ompK* | 55 | 106 | 56 | recombination |
| 3 | 11 | *mdh* | 2 | 31 | 1 | point mutation |
| 3 | 88 | *recA* | 2 | 41 | 2 | recombination |
| 3 | 13 | *recA* | 2 | 3 | 1 | recombination |
| 3 | 91 | *recA* | 2 | 44 | 3 | recombination |
| 9 | 25 | *recA* | 2 | 5 | 51 | recombination |
| 9 | 56 | *recA* | 2 | 20 | 46 | recombination |
| 9 | 57 | *recA* | 2 | 21 | 47 | recombination |
| 68 | 70 | *recA* | 24 | 25 | 24 | recombination |
| 68 | 34 | *recA* | 24 | 9 | 12 | recombination |
| 68 | 173 | *recA* | 24 | 26 | 25 | recombination |
| 104 | 66 | *recA* | 54 | 23 | 6 | recombination |
| 104 | 29 | *recA* | 54 | 6 | 18 | recombination |
| 104 | 105 | *ompK* | 59 | 103 | 1 | point mutation |
| 112 | 135 | *recA* | 60 | 79 | 13 | recombination |
| 112 | 149 | *recA* | 60 | 91 | 1 | point mutation |
| 13 | 14 | *ompK* | 55 | 78 | 56 | recombination |
| 13 | 15 | *ompK* | 55 | 91 | 1 | recombination |
| 13 | 16 | *mdh* | 2 | 4 | 1 | recombination |
| 13 | 17 | *mdh* | 2 | 22 | 1 | recombination |
| 14 | 18 | *mdh* | 2 | 22 | 1 | recombination |
| 19 | 12 | *ompK* | 55 | 77 | 1 | recombination |
| 19 | 21 | *ompK* | 55 | 91 | 1 | recombination |
| 4 | 36 | *recA* | 2 | 11 | 51 | recombination |
| 5 | 79 | *recA* | 2 | 34 | 36 | recombination |
| 135 | 134 | *ompK* | 36 | 24 | 66 | recombination |
| 135 | 149 | *recA* | 79 | 91 | 1 | point mutation |
| 135 | 112 | *recA* | 79 | 60 | 1 | recombination |
| 112 | 111 | *ompK* | 36 | 1 | 57 | recombination |

**Table 8** Variant alleles within each of the nt-SLVs.

Within *ompK*, 33 of the putative recombination events are nonsynonymous substitutions. Therefore, although *recA* is undergoing a similarly high rate of recombination to *ompK*, the functional relevance of recombination may be greater at *ompK* as more of the changes introduced by recombination are nonsynonymous. Similarly, although in terms of allelic change, the contribution of recombination to clonal diversification is almost double in *recA* than in *ompK*, it is notable that all but 6 of the recombined alleles in *ompK* correspond to at least 54 nucleotide changes (mean = 23.8), whereas the mean number of nucleotide changes introduced per allele into *recA* is much lower (n = 17, mean = 20.4).

### 6.3.9 Splits decomposition analysis

Splits decomposition analysis was applied to the nucleotide clonal complex SLV data to test for evidence of recombination (Huson and Bryant, 2006). Simple bifurcating tree-like graphs are observed with nt-CC9, nt-CC68, nt-CC104 and nt-CC172, each with splittability indexes of 100 % (data not shown). This observation is expected with minor clonal complexes, as they most likely have recently evolved, have yet had time to diversify and have limited recombination. Surprisingly, nt-CC135, the minor complex composed exclusively of NZ isolates, forms a single reticulation, central to the graph (Fig. 17a). This is due to 2 large scale recombination events within *ompK*. Significant reticulate structure is observed within nt-CC3 (Fig. 17b), with network structures present at both the base of the graph and at a single edge, indicating extensive recombination has taken place within this complex. STs presented on the long branches, with the exception of ST18, are more diverse members of the clonal complex having undergone considerable recombination events (> 36 bp). nt-CC3 is composed entirely of UK isolates and has a splittability index of 92.87 %, a high goodness of fit for the data. Interestingly, nt-CC172, the only clonal complex presented here which contains isolates from both UK and NZ, does not show any reticulate structure. Instead reticulate structure is observed within the UK (nt-CC3) and NZ (nt-CC135).

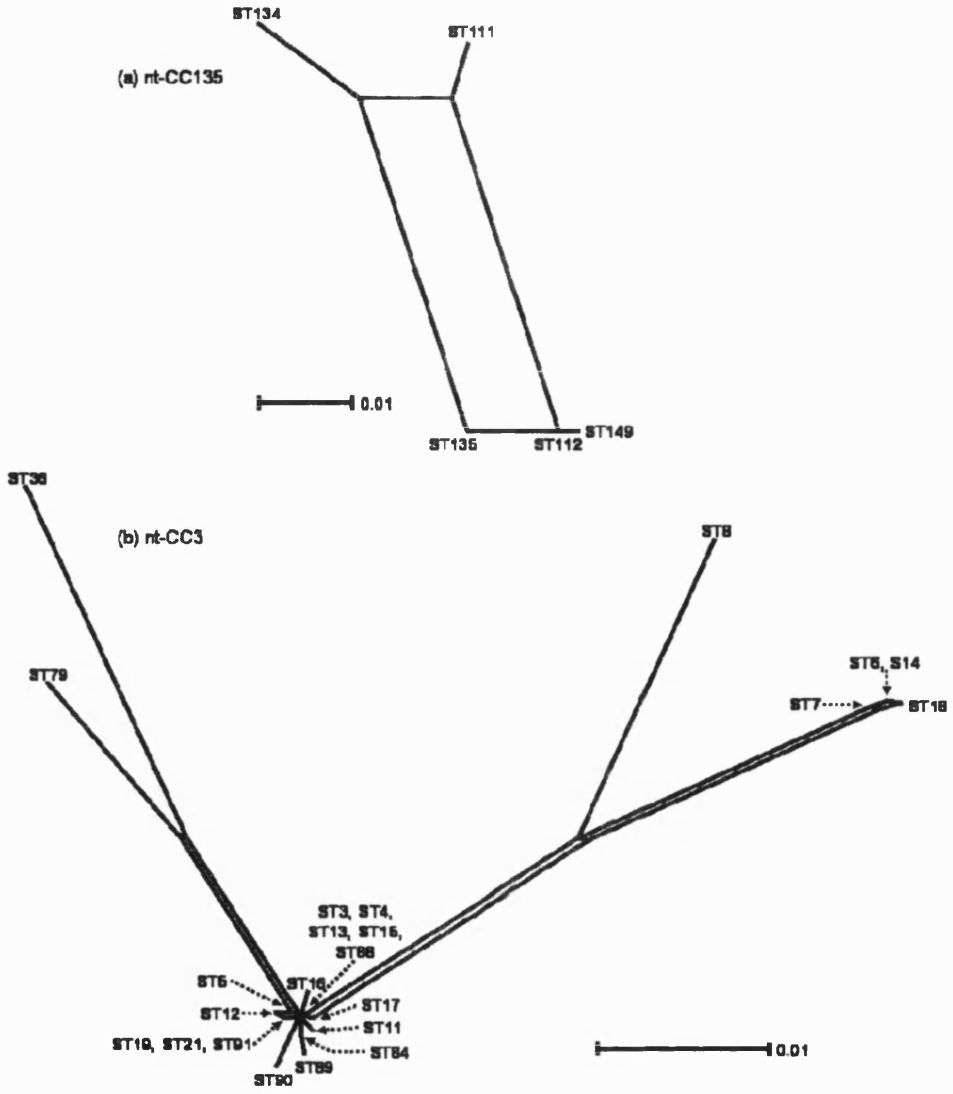



**Figure 17** Splits decomposition analysis of two clonal complexes, a) nt-CC135 and b) nt-CC3, present within the *Vibrio* UK and NZ population

6.3.10 Differentiation between UK and NZ populations

The classification index (CI) as described by Jolley *et al.*, (2005) was calculated for comparisons of the UK and NZ isolates locus haplotypes (Table 9). This analysis confirms that the samples are statistically distinct ($P<0.001$) on the basis of the frequencies of the sequences of all three loci, which is consistent with the phylogenetic and clustering analyses described above. This strong differentiation is

consistent with low rates of migration or the local expansion of specific genotypes, possibly owing to local adaptation.

| Loci | CI |
|---|---|
| *recA* | $P < 0.001$ |
| *mdh* | $P < 0.001$ |
| *ompK* | $P < 0.001$ |

**Table 9** Differentiation between the UK and NZ Vibrio populations.

6.3.11 Further evidence for recombination within the sequences

A further two tests were used to examine the extent of recombination within the population. The first is the coalescence approach of Fearnhead and Donnelly (2001). This approach is used to calculate the population-scaled recombination parameter ρ, and has been widely used to examine MLST datasets (Perez-Losada *et al.*, 2006). This parameter was estimated for all three loci separately for the two UK and NZ samples in turn (Table 10).

| Loci | | | | | |
|---|---|---|---|---|---|
| *recA* | | *mdh* | | *ompK* | |
| UK | NZ | UK | NZ | UK | NZ |
| 5.051 | 20.202 | 18.1 | 24.242 | 26.263 | 100 |

**Table 10** Estimations of the recombinational parameter ρ within the UK and NZ *Vibrio* populations using LDhat.

The estimated values of ρ range from 5.051 (*recA* in the UK sample) to 100 (the maximum value, observed for *ompK* in the NZ sample). For all three genes, the value is greater for the NZ than the UK population, which probably reflects the greater diversity in the NZ sample, and the highest value is observed for *ompK*, as expected. Interestingly, this test suggests a higher impact of recombination within *mdh* than *recA* for both the NZ and UK samples, which is not consistent with the

phylogenetic or SLV analyses described above. Whereas the latter two tests primarily detect changes affecting the whole gene, LDhat is more sensitive to very small replacements effecting the distribution of polymorphisms within the gene sequence and this may help to explain this discrepancy.

The second test used was Sawyer's Runs Tests (Sawyer, 1989). This test has been widely used to infer recombination events within aligned sequences by determining if regions of sequence pairs have more consecutive identical polymorphic sites in common than would be expected by chance. The SSCF (sum of the squares of the condensed fragments) is calculated from silent polymorphic sites only, whereas the SSUF (sum of the squares of uncondensed fragments) also considers monomorphic sites in calculating the significance of "runs" of polymorphic sites between sequences. A gene conversion event tends to increase the values of SSCF and SSUF, as it results in an identical region within two sequences and may produce an unusually long fragment. The SSUF is more likely to detect larger gene conversions than the SSCF, as this takes all the invariant sites into account. Significance is gauged by randomly permuting the sequence 1000 times whilst retaining the frequencies of the polymorphisms in the input data. The results of this test are shown in Table 11.

| | Loci | | | | | |
|---|---|---|---|---|---|---|
| Method of | *recA* | | *mdh* | | *ompK* | |
| analysis | UK | NZ | UK | NZ | UK | NZ |
| SSCF | 0.7391 | 0.6976 | 0.6034 | 0.3995 | 0.14 | 0.0022 |
| (P-value) | (P > 0.05) | (P > 0.05) | (P > 0.05) | (P > 0.05) | (P > 0.05) | (P < 0.05) |
| SSUF | 0.8136 | 0.774 | 0.7821 | 0.38 | 0.2302 | 0.0621 |
| (P-value) | (P > 0.05) | (P > 0.05) | (P > 0.05) | (P > 0.05) | (P > 0.05) | (P > 0.05) |

**Table 11** Summary of the recombinational exchange using the Sawyer's Runs Test.

Neither the SSCF or SSUF statistics revealed significant "mosaic" structure (a non-random distribution of polymorphisms) within any of the three genes with the exception of *ompK* within the NZ population ($P < 0.05$). Again, this is inconsistent with previous analyses which have revealed high rates of recombination, particularly

within *ompK* and *recA*. As for LDhat, Sawyer's Runs test examines the rate at which recombination has effected the distribution of polymorphisms within the sequences, and will be insensitive to events which have resulted in the replacement of whole alleles.

## 6.4 DISCUSSION

Microbial biodiversity has become one of the most intensely studied areas of biological research in the 21st century. With an estimated 99% of all microorganisms yet to be discovered, there is an urgent need to understand the evolutionary forces responsible for the emergence and maintenance of this diversity. In particular, two fundamental questions remain unanswered. The first, how many types ("species") of bacteria co-exist together in the same population, and secondly, to what extent are such types of bacteria naturally organised into discrete phylogenetic clusters (Acinas *et al.*, 2004)? Previous studies assessing the ecology of *Vibrio* spp. have focused on specific members of the genus (Jiang *et al.*, 2000b; Wright *et al.*, 1996). However, the diversity and dynamics of co-occurring *Vibrio* populations have been rarely addressed (Acinas *et al.*, 2004; Heidelberg *et al.*, 2002; Thompson *et al.*, 2004c).

Three MLSA genes were used in this study to assess the genomic diversity within the *Vibrio* genus and to estimate the mechanisms involved in this diversity. An extensive dataset of 285 isolates has been generated for three loci (*recA, mdh, ompK*). The isolates were recovered from the Gower Peninsula, UK (n=180) and from two sites on the North Island (East & West coasts), New Zealand (n=105), which are 33 km apart by land and ~1000 km by sea. These data provide the means to address the aims listed in Chapter 1 and these will now be discussed in turn:

The detection of geographic structure using alternative marker loci

i) Species differences

The sampling regime used in this study allows an examination of the degree of geographic structure over two scales, a) localised comparisons between the two sites in NZ, and b) global comparisons between UK and all NZ isolates. The populations were compared in a number of different ways. Firstly, the overall species assignments on the basis of *recA* BLAST hits were used. Although this is a fairly crude tool, it provides a means to get a handle on the different datasets, and also to address the utility of *recA* as a species marker. Comparisons between *recA* and the

other two loci also raise questions concerning precisely what a "species" means in bacteria.

Profound differences were noted in the frequency of each species noted within the UK and NZ samples. Approximately half of the samples from each country consisted of isolates assigned to a single species, *V. cyclitrophicus* in the case of the UK isolates, and *V. pomeroyi* in the case of the NZ isolates. Although one possible explanation is that these two species are locally adapted, it is possible that a large component of the differences between UK and NZ reflect the recent explosion of local clones. It was apparent in Chapter 5, where the isolates from the UK were considered separately, that the composition of the *Vibrio* population changes very dramatically over the year, and that the large *V. cyclitrophicus* clade has resulted from a dramatic local clonal expansion, which all but disappeared from the local environment in the following months. Although both the UK and NZ samples were recovered during the summer months, given this temporal instability on a local scale it is perhaps not surprising that such great differences are noted between the UK and NZ isolates. Larger samples are therefore required in order to tease apart the temporal and geographical dimensions to the observed diversity.

However, the current data does at least provide some tantalising clues. Even given the seasonal effects described above, it is noteworthy that of the 15 species detected, only 5 were noted in both locations. Furthermore, for all but one of these species common to both sites, there was a pronounced bias towards one site or the other, with the minority isolates often being phylogenetically distinct. The exception being *V. kanaloae*, which was more randomly distributed (being represented by 12.8% of the UK isolates and 7.6% of the NZ isolates). Furthermore, there was little phylogenetic distinction between the UK and NZ *V. kanaloae* isolates, suggesting that this represents a genuinely cosmopolitan species and ecological generalist. The overall differences in genotype frequencies between the UK and NZ isolates were confirmed by the $F_{ST}$ test, and found to be statistically significant using the classification index.

On a finer geographical scale, of the 12 species identified in NZ, only five are common to both East and West coast sites. Interestingly, this excludes *V. kanaloae*,

which is only noted on the East Coast, although the numbers of isolates are too small to draw firm conclusions. The most common NZ species, *V. pomeroyi*, is twice as abundant on the East coast than on the West coast. As these two sites are only separated by 33 km by land, environmental parameters such as sea surface temperature and salinity are likely to be similar, thus this the difference in the frequency of *V. pomeroyi* may reflect some local adaptive pressure.

ii) Is *recA* a reliable phylogenetic marker for *Vibrio* spp.?

These data also provide evidence as to the utility of *recA* as a marker for *Vibrio* species assignments. Although the use of 3 gene loci provides less information than the more typical use of 7 loci for MLST studies, in cases where the phylogenetic signal is inconsistent between two loci it does provide a means to deduce which gene is likely to be aberrant by way of comparison with the third gene. Because of this, the use of 3 gene loci should be regarded as the absolute minimum for MLSA studies; without a third gene it would be impossible to tell which of two inconsistent loci is most likely to reflect the true phylogeny. Although the *recA* and *mdh* trees are broadly consistent, several exceptions have been discussed and it is noteworthy that in these cases the *ompK* tree tends to implicate *recA* as the aberrant gene, suggesting that recombination has affected *recA* more than *mdh*. Indeed, it is possible to quantify the reliability of *recA* as a species marker using the UK *V. cyclitrophicus* clade. This presents quite a conservative measure of reliability, as this species appears to represent a recent clonal expansion, hence marker loci would expect to be more reliable in identifying members of this clone compared to an older, less cohesive, species. Nevertheless the data suggest that roughly 5% of the strains assigned as *V. cyclitrophicus* by *recA* in fact belong to other species, and similarly that 5% of the *V. cyclitrophicus* strains in the population would be assigned as other species by *recA*. For studies aimed at quantifying the presence of a given pathogenic species in an aquaculture setting, for example, these misassignments may present a significant source of error.

iii) Alternative markers and nested species

Darwin's fundamental concept of all living organisms fitting into a nested hierarchy remains a cornerstone of evolutionary and systematic studies (Darwin, 1859).

Studies on the phylogenetic clustering of prokaryotic phyla, genera and species are no exception, and it is clear that clustering occurs on many different levels, dependent upon the resolving power of the data. The use of 16S rRNA, for example, is very limited in detecting clusters within species, or between very closely related species, whereas the use of hyper-variable loci reveals clusters indistinguishable on the basis of conserved housekeeping genes. In principal, therefore, it should be straightforward to characterise isolates according the degree of resolution required, and at some threshold level, demarcate the clusters as "species". These data, however, highlight an important problem with this approach, in that high rates of horizontal transfer will mean that not all clusters are monophyletic, and hence the patterns will be non-hierarchical. The isolates assigned as *V. pomeroyi* in the current data provide a striking example. For *recA*, all *V. pomeroyi* isolates cluster together to the exclusion of other species, and share a common ancestor; indeed this has to be the case given that species are assigned on the basis of this gene. At *mdh*, the cohesion and monophyly of this species is more questionable. First, the strains are interspersed with a cluster assigned as *V. splendidus*, second, a cluster of *V. pomeroyi* appears more closely related to *V. kanaloae* than to other *V. pomeroyi* clusters, and third, a distinct cluster of *V. pomeroyi* appears to fall on a separate branch than the others.

At *ompK*, the assignment of *V. pomeroyi* as a distinct species becomes essentially meaningless, as these strains correspond to multiple clusters dispersed throughout the tree. This illustrates that a second layer of pragmatism is required in order to develop a working definition of species. Until now, the emphasis has focused on "where to draw the line" – in other words, what is the required level of sequence similarity before two strains can be assigned as conspecifics (for example, the 97% cut-off in 16S rRNA sequences, or 70% in DNA-DNA hybridisation, as discussed in Chapter 1). However, these data suggest it is also necessary to define a fraction of the genome which will correspond to the cluster (e.g. 98% identity at 95% of genes).

Evidence for recombination

Although there is evidence to support clonal expansion of the *Vibrio* population, frequent and diverse recombination events are observed within both the UK and NZ

populations. Calculation of the R/M parameter, Sawyer's Runs Test and the population-scaled recombination rate help to clarify the nature of the recombination events observed, that they have generally involved the transfer of whole genes, and there is much more limited evidence to suggest that "blocks" of sequences have been replaced within the boundaries of the sequenced alleles. In addition, the recombinational replacements may be fairly recent, and have yet to be subjected to further recombination/point mutation events. Furthermore, this may reflect the potential mechanism of recombination. Transformation generally results in smaller (more localised) replacements than transduction. As the current evidence within the study suggests that most replacements are large enough to encapsulate the whole gene, this suggests a larger role for transduction (phages) as a vehicle for recombination. Complete genome sequencing has revealed the presence of temperate phage within *Vibrio* genomes (Schoolnik and Yildiz, 2000; Thompson *et al.*, 2004b), and phage are likely to play a key role in HGT. The *ompK* gene encodes a receptor for the broad host-range Vibriophage KVP40, and provides strong evidence that interspecies recombination has been a major driving force in the evolution of this gene. It is likely that the role of transformation may be more limited in the marine, rather than the terrestrial, environment, due to free DNA being present at very low concentrations in sea water. In contrast, marine phages are known to be extremely common, and perhaps represent the biggest reservoir of biodiversity on the planet.

Is everything everywhere?

The data collected from both the UK and NZ suggest the hypothesis of Bass-Becking and Beijerrinck is true, that is, "*Everything is everywhere, but the environment selects*" (Beijerinck, 1913). Differences between the two sites suggest limited migration of species between the UK and NZ, and that the populations are influenced by local adaptation. This is most evident within *ompK* which provides clear clustering of isolates according to their country of origin. At a local scale within both the UK *Vibrio* populations, little difference was observed. This is true also for the two populations which were sampled at both East and West coast sites (data not shown), and even on a larger scale with both the overall East and West coast NZ populations. Of the 172 genotypes included in this study, none are present in both UK and NZ populations, suggesting that "cosmopolitan species" do not exist.

However, the two most evolutionary conservative loci used in this study, *recA* and *mdh*, each possess allelic variants which are present in both the UK and NZ *Vibrio* population (*recA*, n=3 and *mdh*, n=4), which is evidence of extensive migration. Instead, the term "cosmopolitan allele" should be used to describe globally distributed alleles such as these.

# CHAPTER SEVEN

# DEVELOPMENT OF AN INSECT MODEL OF INFECTION BY THE IMPORTANT HUMAN PATHOGEN *STAPHYLOCOCCUS AUREUS*

## 7.1 INTRODUCTION

### 7.1.1 *Staphylococcus aureus*

*Staphylococcus aureus* is an opportunistic pathogen which causes a wide range of symptoms in man, including bone and joint infections and endocarditis (Kluytmans *et al.*, 1997). *S. aureus* has become a global public health burden due to the rise of multi-drug resistant strains (e.g. MRSA, VRSA: meticillin- and vancomycin-resistant *Staphylococcus aureus*), and is a major cause of community-acquired and nosocomial sepsis. Despite its pathogenicity, the bacterium is carried asymptomatically in the anterior nares by approximately 20 % of the UK population. Large proportions (±60 %) of individuals carry *S. aureus* intermittently, characterised by shifts in the frequencies of strains over time (Kluytmans *et al.*, 1997; Kiser *et al.*, 1999). Despite this high carriage rate the frequency of invasive disease within the community is <0.02 % in the UK population. Rates of nasal carriage are known to be higher within diabetics, health care workers and drug users (Kluytmans *et al.*, 1997; Peacock *et al.*, 2002).

### 7.1.2 Bacterial genetic factors and their influence in invasive disease

It is strongly suspected that differences in host and bacterial genetics influence the probability that colonization of *S. aureus* will lead to invasive disease, although in general it has proved difficult to pin-point the role of specific genetic factors. Single bacterial virulence determinants have occasionally been associated with specific disease symptoms (e.g. TSST and toxic shock syndrome), although disease aetiology in general is typically thought to be determined by a complex interplay of multiple bacterial (and possibly host) genes.

Peacock *et al.* used PCR assays to examine the presence or absence of 33 virulence determinants within 334 natural isolates of *S. aureus* recovered from asymptomatic carriage (n = 179) and invasive disease (n = 155) (Peacock *et al.*, 2002). 7 out of the 33 virulence genes (*fnbA, cna, sdrE, sej, eta, hlg, & ica*) were significantly more likely to be present in strains from invasive disease than in strains from asymptomatic carriage. The 7 genes

had a cumulative effect, resulting in an apparently linear relationship between the number of virulence determinants carried by a particular isolate and the chance that this isolate was recovered from a patient with disease ($P < 0.0001$; Figure 1).

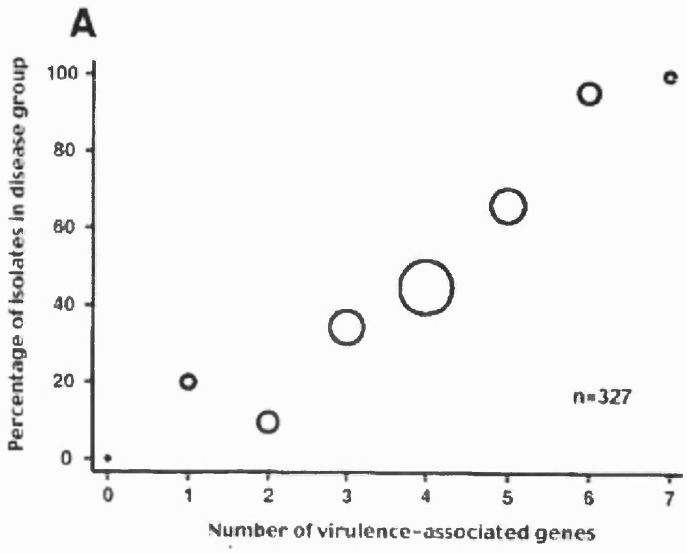



**Figure 1** Relationship between the number of virulence-associated genes and the proportion of isolates from cases of disease. For example, 20 % of isolates with one virulence-associated gene were from cases of disease and 80 % were carriage isolates. The area of each circle is proportional to the number of isolates with that number of determinants (ranging from zero to seven genes), with n = 2, 15, 21, 64, 140, 56, 22, and 7, respectively. Figure from Peacock *et al.*, (2002).

7.1.3 *S. aureus* virulence

The term virulence refers to the quantitative ability of a microorganism to cause disease. *S. aureus* pathogenicity is multi-factorial and is largely dependent on the expression of a wide range of virulence determinants, which influence the ability of strains to cause invasive or innocuous disease; this is a reflection of the various virulence determinants each strain possesses (Jarraud *et al.*, 2002; Feil *et al.*, 2003; Novick, 2003). Hundreds of putative virulence loci have been identified in the *S. aureus* genomes, and several global

regulatory loci exist (*agr, sae, sar*) that play a key role in regulating the expression of these virulence factors (Giraudo *et al.,* 2003; Novick, 2003; Jarraud *et al.,* 2002).

The accessory gene regulator (*agr*) locus regulates the production of most staphylococcal exoproteins, including exoenzymes, toxins, surface proteins, and other virulence factors, via a two-component signal transduction system, by means of down-regulation of surface proteins and up-regulation of secreted proteins during in vitro growth (Novick *et al.*, 1993). The *agr* locus consists of two divergently transcribed operons, driven by promoters P2 and P3. The P2 operon consists of four genes; *agrB*, *agrC*, *agrD* and *agrA* (Novick *et al.*, 1993). In summary, AgrB activity leads to secretion of the autoinducing pheromone, AgrD, which binds to and activates the histidine kinase receptor, AgrC, which subsequently activates the response regulator, AgrA (Novick *et al.*, 1993; Robinson *et al.*, 2005). The P3 operon consists of RNAIII, the regulatory effector molecule of the *agr* locus, and the gene encoding δ-haemolysin, *hld* (Jarraud *et al.*, 2000; Novick *et al.*, 1993). The P2 P3 signalling pathway is activated by the ligand AIP (*agr*-encoded autoinducing peptide). A polymorphism in the amino acid sequence of the AIP and of its corresponding receptor divides *S. aureus* isolates into four major groups; Agr type I, II, III and IV. Within a given group, each strain produces a peptide that can activate the *agr* response in the other strains of the same group, whereas the AIPs belonging to different groups cause the inhibition of the *agr* response (Jarraud *et al.*, 2002). This type of bacterial interference is thought to influence colonisation dynamics, by enhancing or inhibiting the ability of *S. aureus* to colonise in the presence of resident strains (Ji *et al.*, 1997). Furthermore, an association between *agr* type and strains causing disease has been described. For example, *agr* type III has been associated with toxic shock syndrome (Francois *et al.*, 2006) and *agr* type IV has been associated with scalded skin syndrome (Jarraud *et al.*, 2000; Jarraud *et al.*, 2002).

The *sar* locus produces SarA, a DNA binding protein, which acts in an *agr*-independent manner (Giraudo *et al.,* 2003). SarA acts on the expression of virulence factors by indirectly upregulating RNAIII transcription and stimulating/repressing the transcription of other virulence genes such as *fnbA, cna* and *hla.*

The *sae* locus encodes *saeR* and *saeS*, a response regulator and a histidine protein kinase respectively, which upregulate virulence factors such as DNase and haemolysins at the transcriptional level.

### 7.1.4 The use of invertebrates as model systems

Much of what is known concerning the role of specific virulence factors in *S. aureus* is derived from using animal models, particularly mouse, rat and rabbit. However, there has recently been increasing interest in the development of invertebrate models of infection. As well as being cheaper and potentially far simpler in terms of experimentation, this would also help to alleviate the ethical issues of using animal models. Therefore, it is necessary to ask to what extent invertebrates are realistic model systems to use for human pathogens, and how do their immune responses compare with those of higher organism?

The ubiquitous threat posed by pathogens, bacterial and otherwise, has resulted in the evolution of a wide variety of defensive strategies throughout the eukaryotes. Immunity – the state of protection from infectious diseases – has both non-specific and specific components in vertebrates. The non-specific component, innate immunity, refers to disease resistance mechanisms effective against a broad range of pathogens, whereas the specific component (adaptive immunity) relies upon antibody recognition and is highly pathogen-specific (Goldsby *et al.,* 2000). The innate immune response is essentially instantaneous and provides the first line of defence during the critical period after the host's exposure to a pathogen, whereas the adaptive immune response typically occurs within 4 – 6 days after the initial exposure to an antigen. Adaptive and innate immunity in mammals do not operate independently of each other, but instead function as a highly interactive and co-operative system thus being more effective than either could be alone (Kimbrell & Beutler, 2001). Insects lack an acquired immune system, but instead have a well-developed innate immune response. They are constantly exposed to potential pathogens, but only a few will go on to develop an infection.

The current model of insect innate immunity suggests a number of interconnected reactions form this defence. The primary defences include the integument and gut, which act as physical barriers to infection, whereas the cellular defence reactions are mediated by haemocytes, including phagocytosis, nodulation and encapsulation (Table 1). Finally the humoral defence reactions are responsible for the synthesis of antimicrobial peptides and proteins primarily by the fat body and midgut, and the regulation of coagulation and melanization of haemolymph (Gillespie *et al.,* 1997; Lavine & Strand, 2002). The insect innate immune system uses a variety of germline-encoded pattern-recognition receptors (PRRs) that recognize conserved microbial structures or pathogen-associated molecular patterns (PAMPs) produced by microorganisms like bacteria and fungi. PRRs can be involved in both the humoral and cellular innate immune responses (Lavine & Strand, 2002; Yu *et al.,* 2002; Mushegian & Medzhitov, 2001).

| Reaction | Function |
|---|---|
| Phagocytosis | Haemocyte engulfment of biotic targets (e.g. bacteria, yeast, apoptotic). With *M. sexta*, granular & plasmatocytes are type of haemocyte that are phagocytic. |
| Nodulation | Multiple haemocytes bind to aggregates of bacteria forming a sheath of haemocytes around target. |
| Encapsulation | Multiple haemocytes bind to larger targets (e.g. parasitoids & nematodes) forming a sheath of haemocytes around target. |

**Table 1** Functions of the cellular defence reactions

Innate immunity developed before the separation of vertebrates and invertebrates millions of years ago. Consequently, many of the signal pathways and pathogen-receptor molecules are conserved between insects and humans. For example, with the mammalian innate immune response, NF-κB is activated in the presence of a pathogen-generated signal, mediated by IL-1 or TOLL receptors. In adult Drosophila melanogaster, TOLL receptors

signal via Dorsal, the NF-κB homologue, to activate the immune response (Mylonakis and Aballay, 2005; Takeda and Akira, 2003). Additionally, mammals and insects produce a similar class of antimicrobial peptides called defensins, in a pathogen inducible manner (Hoffmann *et al.*, 1999; Lehrer and Ganz, 1999). Because of these similarities, insect model systems for the study of infection are advantageous in the study of virulence mechanisms with pathogenic bacteria, potentially even those that are considered human pathogens. Already a number of invertebrate model systems have recently been described and are under development for bacterial virulence including *Caenorhabditis elegans* (Sifri *et al.*, 2003), *Galleria mellonella* (Jander *et al.*, 2000) and *Bombyx mori* (Kaito *et al.*, 2002).

*Manduca sexta* larvae have a number of advantages as an *in vivo* model of infection. Firstly, this host possesses an innate immune system similar to that of higher organisms. Larvae have practical attributes such as a short generation time (from eggs to fifth instar larvae takes approximately 17 days), a short lifespan and they are easily maintainable in the laboratory. However, one of the most important advantages is the size of *M. sexta* larvae, which is appropriate for handling during injection of bacteria with syringes; this is a practical problem in some smaller sized insects we know more about genetically, e.g. *Drosophila melanogaster*. Additionally, the use of insect models allows the study of pathogens at human body temperature, and overcomes the financial and ethical constraints imposed with mammalian models.

## 7.2 METHODS

### 7.2 1 Preparation of *M. sexta* diet

*M. sexta* were maintained on a nutrient supplement composed of various ingredients. In preparation, a "Premix" was initially prepared containing 540g deactivated dried yeast (DCL Yeast Ltd, UK), 2700g wheatgerm (Seasons, UK), 1260g casein (Sigma, USA), 1080g sucrose (Tate & Lyle, UK), 360g Wesson's Salt (Sigma, UK), 36g Choline Chloride (Sigma, UK), 72g Cholesterol (Sigma, UK), 36g methyl paraben (Sigma, UK) and 54g Sorbic acid (Sigma, UK). Once the ingredients are thoroughly mixed together, the Premix should be stored at 4°C until required.

The following protocol should be followed for the preparation of one batch of *M. sexta* diet:

- Boil 650ml of distilled water, and heat and dissolve 22.5g of plant agar (IDG, UK) in 1 L of distilled water; mix thoroughly,
- To the distilled water/agar mix add:
  - 336g Premix
  - 8ml 10% formalin
  - 4ml Linseed oil (J.M. Loveridge Plc, UK)
  - 4ml corn oil (Bestfoods UK Ltd, UK)

  Mix for 20 minutes, and allow the mixture to cool to below 50°C,
- Once the temperature is < 50°C, add to the mixture:
  - 16g Ascorbic acid (Amresco, USA)
  - 0.2g Vanderssant vitamins (ICN Biomedicals Inc, USA) dissolved in 50ml distilled water Mix thoroughly and pour into a metal tray lined with aluminium foil and allow to set. Store at 4°C.

### 7.2.2 *M. sexta* feeding and handling

*M. sexta* first day first instar larvae were reared on artificial diet for approximately 13 days in small plastic containers at 24°C. At day one of the fifth stadium larvae were transferred to larger containers with fresh diet. All unused *M. sexta* larvae and end-point larvae were terminated at -20°C.

### 7.2.3 Bacterial strains

*Staphylococcus aureus* strains used in this study have been isolated from Oxfordshire, United Kingdom. These strains were isolated over a two year period from the anterior nares of asymptomatic blood donor and from patients with disease in hospital, and in the community. These have been previously characterised by MLST (Feil *et al.*, 2003). The following prefixes are used to distinguish the source of the *S. aureus* strain; D = donor; C = community acquired disease; H = hospital acquired disease.

Five pairs of natural *S. aureus* isolates were included in this study (Table 2). Within each pair is a low virulence isolate, which either contains zero of the selected seven virulence genes, or one of the seven virulence genes (*hlg*); these strains have been recovered from asymptomatic carriage. The remaining strain is a high virulence isolate, which contains all 7 of the selected virulence genes; these are the strains which have been recovered from invasive disease.

| Low virulence | | High virulence | |
|---|---|---|---|
| D339 | (0/7) | C900 | (7/7) |
| D341 | (0/7) | C640 | (7/7) |
| D219 | (0/7) | C720 | (7/7) |
| D334 | (1/7) | H182 | (7/7) |
| D412 | (1/7) | H617 | (7/7) |

**Table 2** Isolates used for the comparison of low and high virulence *S. aureus* infection

### 7.2.4 Preparation and storage of cell and DNA stocks

Depending on culture requirements, *S. aureus* strains were either grown on TS agar (Oxoid, UK) with 5 % w/v defibrinated sheep's blood (TCS Microbiology, UK), or TS broth (Oxoid, UK) for a minimum of 12 hours at 37°C. For Agr competition assays, strains were grown overnight at 37°C in 5ml of brain heart infusion broth (Oxoid, UK).

Several colonies were resuspended in 1ml TS broth (Oxoid, UK) with 15% w/v glycerol and stored at -80°C providing frozen cell stocks.

Further colonies were resuspended in 400µl of lysis solution containing 20µl Lysostaphin (500units $ml^{-1}$), 20µl Lysozyme (5000units $ml^{-1}$), 360µl TE Buffer (0.2ml 0.5M EDTA ph8.0, 1ml Tris ph8.0, 98.8ml $dH_2O$).

Genomic DNA was extracted using the DNeasy tissue kit (Quiagen, UK). DNA stocks were then stored at -25°C.

### 7.2.5 Injection of *S. aureus* into haemolymph of *M. sexta*

Fifth instar larvae were injected with 10µl of a suspension containing $10^4$ washed *S. aureus* cells, unless otherwise stated. Injections were performed directly into the haemocoel at the anterior abdominal segment and pressure was immediately applied at the site of injection for approximately 10 seconds to prevent bleeding. Control larvae were injected with 10µl of sterile Tryptic Soya Broth (Sigma, USA). Syringes (1ml) and needles (22s/2"/2) were obtained from Hamilton, Switzerland. The larvae were maintained at 37°C and used for further analysis.

### 7.2.6 Quantification of *S. aureus* in *M. sexta* larvae

*M. sexta* were immersed in 70 % ethanol, flamed and dissected along the segmental groves in the lid of a petri dish. The contents were transferred to a 50 ml sterile tube (Sterilin, UK) and vortexed for 2 minutes. The carcass was diluted in TSB in 10-fold serial dilutions and spread on 5 % blood agar plates, which were incubated at 37°C. Colony counts were performed and the number of cells in the original sample calculated.

### 7.2.7 Passaging of *S. aureus*

The following methods were used for the passaging of *S. aureus* in *vivo*;

Passaging protocol 1:

- *S. aureus* strain C900 was injected into larvae (3 x n=5) at dose $10^3$ CFU to ensure the full effect of virulence could be achieved, but without premature larval death; this may have occurred at $10^4$ CFU. Larvae were incubated at 37°C.
- Day 3 post-injection, larvae were dissected, homogenised and the homogenate injected at a dose of $10^3$ CFU (calculated from the known number of bacteria recovered after 3 days and diluted accordingly) into "new" larvae. This is a single passage stage. Bacterial stocks were taken for future reference.
- Passaging was repeated every 3 days for a period of 27 days.

Passaging protocol 2:

- *S. aureus* strain C900 was injected into larvae (3 x n=5) at dose $10^3$ CFU. Larvae were incubated at 37°C.
- Day 3 post-injection, larvae were dissected, homogenised, the homogenate diluted in 10-fold serial dilutions, spread plated on 5 % blood agar and incubated at 37°C. The homogenate was stored at 4°C.
- Approximately 12 – 18 hours post-homogenisation, colony counts were performed on the blood agar plates and the relevant serial dilution which produced $10^3$ CFU was noted. The corresponding homogenate stored overnight at 4°C was injected into a "new" larva. Bacterial stocks were taken for future reference.

### 7.2.8 Agr interference between clinical *S. aureus* strains

Forty-five randomly selected strains from the collection described in Feil *et al.* were competed against an individual, marked *S. aureus* strain to measure their relative fitness (Feil *et al.*, 2003). Details of the isolates used can be found in Appendix E. Associations between relative strain fitness and specific genetic factors were subsequently determined. A naturally tetracycline-resistant strain was identified from the collection and is hereafter termed the marked strain. The haemocoel of fifth-instar *M. sexta* larvae were injected with 10µl of media containing equal number of the two competing strains at a density of $10^4$

CFU per insect. The starting densities were measured by plating on TSA agar with and without tetracycline (2 μg/ml) prior to inoculation. The insects were then incubated at 37°C for 3 days, homogenized, and the homogenate serially diluted and plated on TSA agar with and without tetracycline; any insects that died before this final time point were discarded. After 24hrs incubation at 37°C the number of marked (tetracycline resistant) and tester (tetracycline sensitive) colonies was enumerated. Relative fitness was calculated from the ratio of the marked:tester strain Malthusian parameters ($m=ln(N_f/N_o)$ where $N_f$ and $N_o$ are the final and starting densities respectively) (Lenski *et al.*, 1991). The competition experiments were performed in duplicate and the results averaged.

### 7.2.9 Choice of genes, primer design & PCR conditions

Seven housekeeping loci (*arc, aroE, glpF, gmk, pta, tpi, yqiL*) were amplified according to the protocol outlined for the *S. aureus* MLST scheme (Enright *et al.*, 2000). All primer sequences can be found in Appendix E.

All PCR reactions contain the following reagents:

| | |
|---|---|
| 12.5 μl | ReddyMix™ PCR Master Mix (ABgene, UK) |
| | 1.25 u Thermoprime Plus DNA Polymerase |
| | 75 mM Tris-HCl (pH 8.8 at 25°C) |
| | 20 mM $(NH4)_2SO_4$ |
| | 1.5 mM $MgCl_2$ |
| | 0.01 % (v/v) Tween® 20 |
| | 0.2 mM each of dATP, dCTP, dGTP, dTTP |
| | Precipitant and red dye for electrophoresis |
| 9.5 μl | distilled $H_2O$ |
| 1 μl | Forward primer (10 pmol-1, MWG Biotech, Germany) |
| 1 μl | Reverse primer (10 pmol-1, MWG Biotech, Germany) |
| 1 μl | DNA |

The following PCR conditions were used for each of the MLST primers:

| | | | |
|---|---|---|---|
| Initial denaturation | | 3 mins | 94°C |
| 34 cycles: | denaturation | 1 min | 94°C |
| | annealing | 1 min | 55°C |
| | extension | 30 sec | 72°C |
| Final extension | | 10 mins | 72°C |

7.2.10 Methods of Analysis

-Sequence editing and alignment

As described in Chapter 2 Section 2.5 *Nucleotide Sequence Analysis.*

-Relative fitness index

To determine the virulence of *S. aureus* strains during the course of an experiment, the relative fitness index fitness of the larvae was calculated:

relative fitness index = (maximum weight – starting weight) / starting weight.

## 7.3 RESULTS

### 7.3.1 Normal larval development is observed when *M. sexta* are injected with sterile broth

To monitor the normal development of *M. sexta* larvae, four first day fifth instar larvae were injected with 10µl sterile TSB (Fig. 3a). All larvae developed through the fifth instar larval stage until approximately 60 hours post-injection, when they progressed to the wandering stage. This stage was marked by a reduction in body weight (expulsion of gut contents), loss of body moisture and a prominent dorsal heart along the larva's back. This patterns closely mirrors that observed in the control caterpillars which were not injected (Fig. 3b), suggesting that injection with sterile TSB does not significantly effect the growth and development of the larvae.

### 7.3.2 *S. aureus* kills *M. sexta*

A pilot study conducted by Daniel Bridges concluded that a number of human bacterial pathogens were also pathogenic to *M. sexta* (Bridges, 2002). To determine whether the ability to kill *M. sexta* is a common property of *S. aureus*, larvae were injected with either $10^3$ or$10^4$ CFU with strains D339 or C900 (0/7 and 7/7 virulence determinants). Noticeable differences were observed with *M. sexta* development (Fig. 4 & Table 2). Between 0 – 30 hours, no *M. sexta* mortality or wandering stage larvae were observed with either of the two strains. At 48 hours post-injection, the first larval deaths occur with strain C900 at $10^4$ CFU (n=3) and C900 at $10^3$ CFU (n=1), followed by D339 at $10^4$ CFU at 73.5 hours post-injection. By 100 hours post-injection, 6 larval deaths were observed with strains D339 and C900 at $10^4$ CFU, and 2 larval deaths with C900 at $10^3$ CFU. Therefore, we can conclude, i) *S. aureus* kills *M. sexta*, ii) larvae injected with C900 have the highest and fastest mortality rate and iii) the $LD_{50}$ is approximately $10^4$ CFU. A 100 % mortality rate is not observed due to one larva in each category surviving the infection and progressing to the wandering stage.

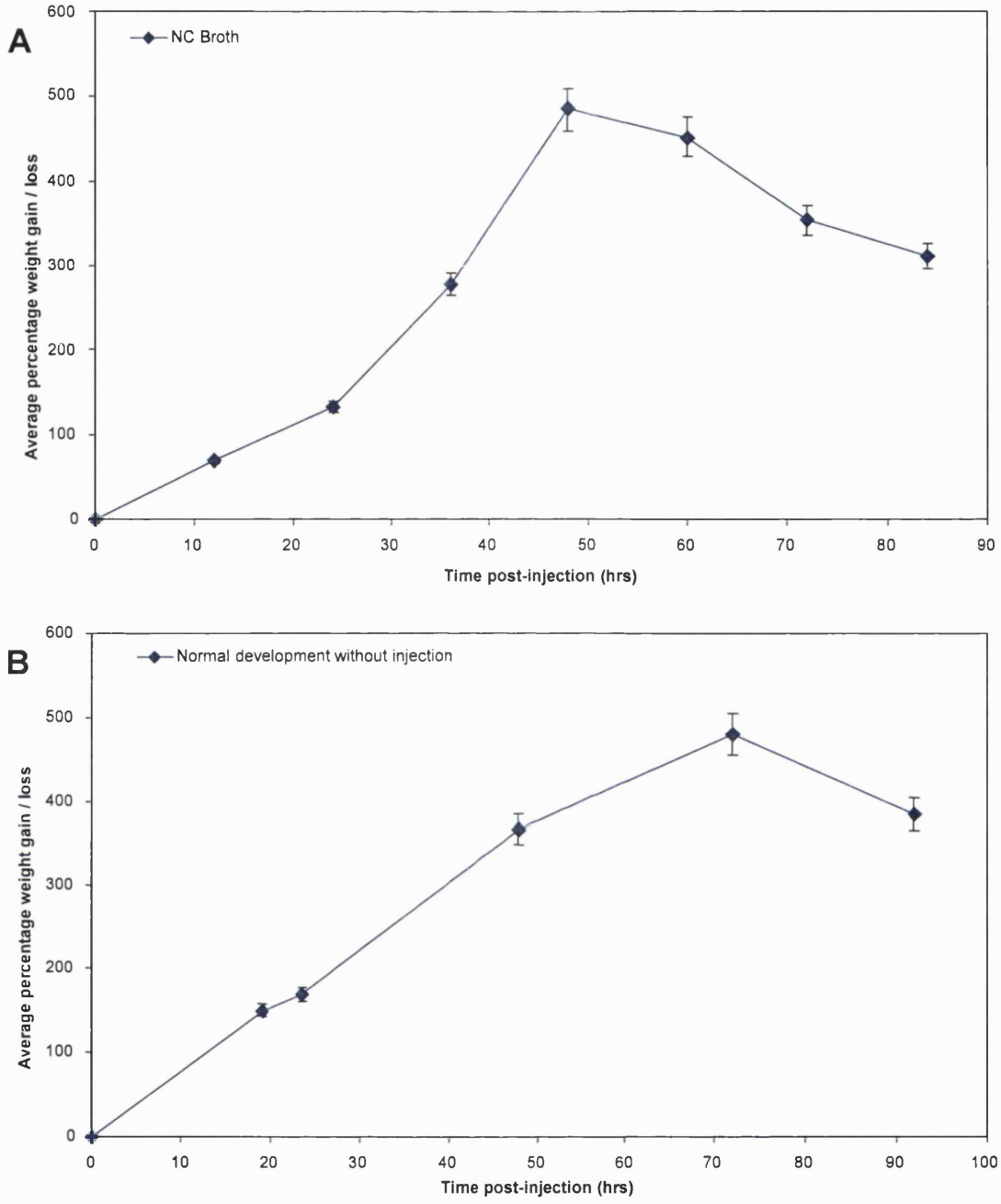



**Figure 3** Graph showing the relationship between weight gain/loss after injection with sterile TSB. The error bars represent 95 % confidence. 1st day 5th instar larvae were a) injected with 10 µl sterile TSB (n=4) and b) allowed to progress naturally through their life-cycle (n=8).

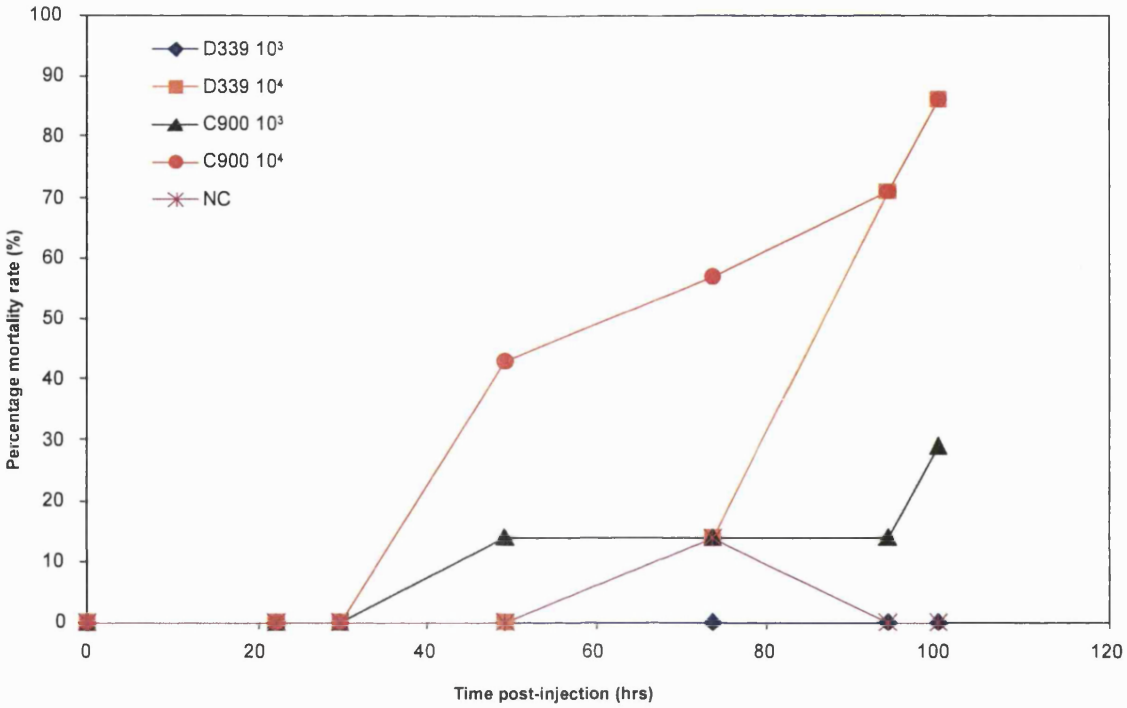



**Figure 4** *M. sexta* mortality following *S. aureus* infection (n = 3-5).

| Strain | Time post-injection (hrs) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 22 | 29.5 | 49 | 73.5 | 94 | 100 |
| D339 $10^3$ | | | | | | | |
| A | 7 | 7 | 7 | 7 | 5 | 0 | 0 |
| D | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| W | 0 | 0 | 0 | 0 | 2 | 7 | 7 |
| D339 $10^4$ | | | | | | | |
| A | 7 | 7 | 7 | 7 | 6 | 1 | 0 |
| D | 0 | 0 | 0 | 0 | 1 | 5 | 6 |
| W | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| C900 $10^3$ | | | | | | | |
| A | 7 | 7 | 7 | 6 | 5 | 3 | 2 |
| D | 0 | 0 | 0 | 1 | 1 | 1 | 2 |
| W | 0 | 0 | 0 | 0 | 1 | 3 | 3 |
| C900 $10^4$ | | | | | | | |
| A | 7 | 7 | 7 | 4 | 2 | 1 | 0 |
| D | 0 | 0 | 0 | 3 | 4 | 5 | 6 |
| W | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

**Table 3** *S. aureus* kills *M. sexta* (n=7). Key: A = alive, D = dead & W = wanderers.

### 7.3.3 Live bacteria are required to kill *M. sexta*

It was necessary to determine whether *Manduca* mortality required the presence of live bacteria or the supernatant fraction (soluble proteins & toxins). Following overnight incubation, isolates D339 and C900 were heated at 121°C and compared to untreated growth from same overnight suspension. Both strains were washed and the supernatant discarded to remove any toxins. Prior to 39 hours post-injection (Fig. 5), no significant differences are observed between heat-treated and untreated cells. However, at 39 hours post-injection clear differences are observed in average percentage weight gain, with the heat-killed cells having a much higher average percentage weight gain (similar to that of the control larvae). Previous research (Bridges, 2002) had shown injection of supernatant did not result in *M. sexta* mortality. This suggests virulence requires the presence of live *S. aureus* and is not solely due to the secretion of toxins and soluble proteins into the supernatant. Overall, a lower percentage weight gain was observed with C900 in comparison to D339 and the negative control consistent with results discussed later in this chapter.

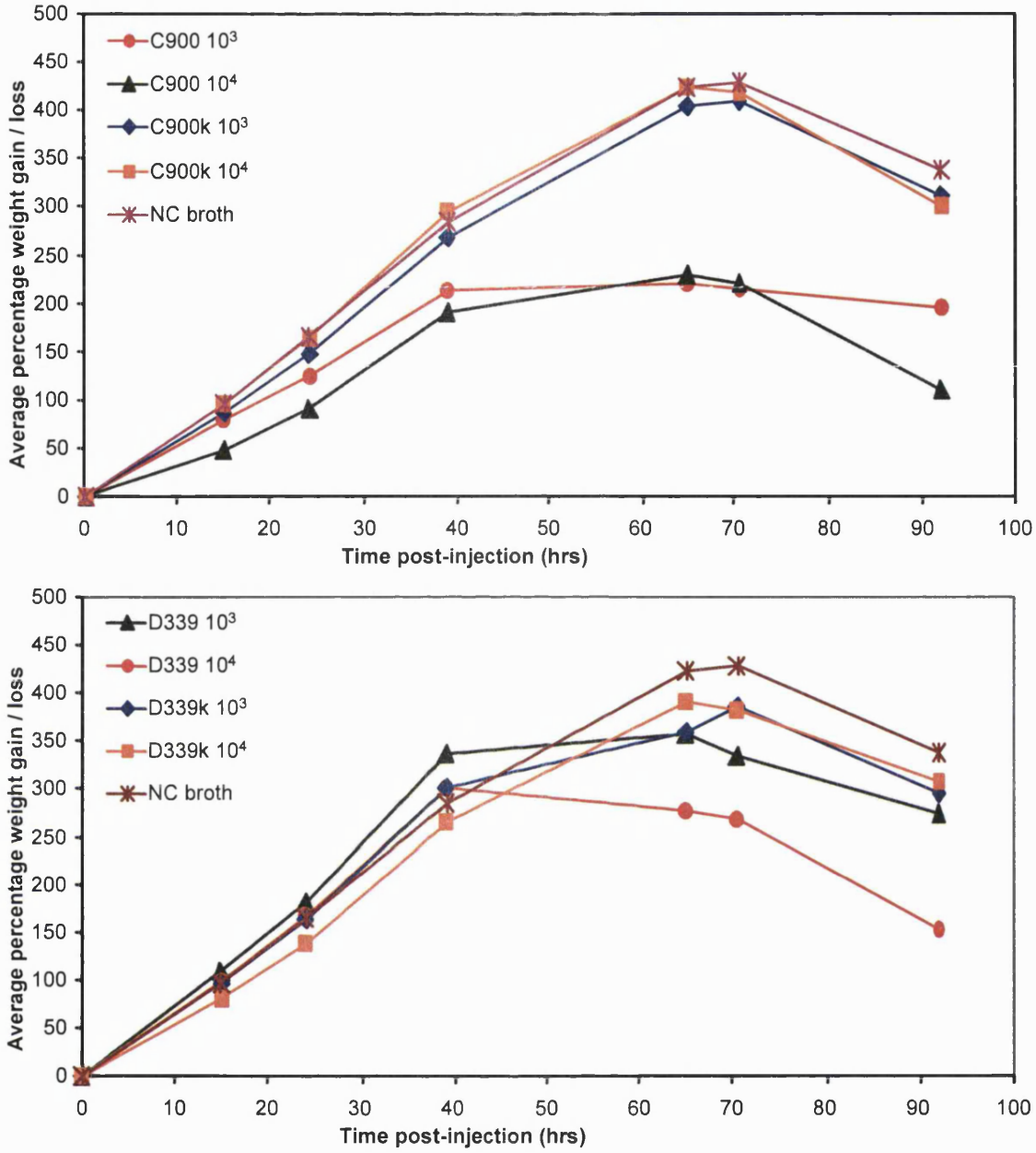



**Figure 5** The presence of live *S aureus* cells is required to kill *M. sexta*. Larvae were injected with 10 µl of either heat-treated or untreated *S. aureus* (n = 4-5). Heat treated strains can be identified by the letter "k" at the end of strain names.

### 7.2.4 *S. aureus* recovery from *M. sexta* following infection

To determine whether *S. aureus* proliferated in the larval bodies, *M. sexta* larvae were injected with C900 at doses $10^3$ and $10^4$ CFU to estimate the number of recoverable bacteria along the course of the infection. *M. sexta* larvae were homogenised according to the protocol outlined in Section 7.2.6 *Quantification of S. aureus in M. sexta larvae.* The number of recoverable bacteria increased by a factor of $10^5$ in 72 hours (Fig. 6), approximately a log order of magnitude every 12 hours, which is comparable to the rate of growth in nutrient media. By 72 hours, larvae injected with C900 $10^4$ CFU showed a maximum proliferation of $10^9$, with strain C900 at dose $10^3$ CFU reaching 7.7 x $10^8$ CFU.

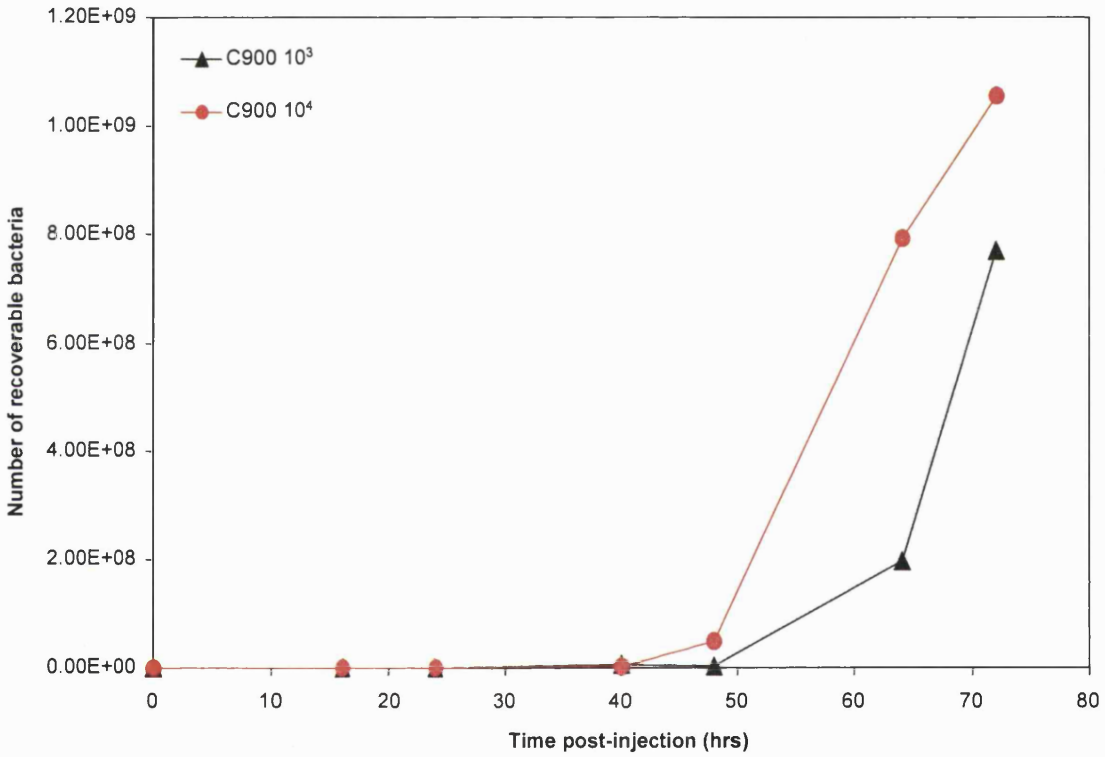



**Figure 6** Graph showing the relationship between *S. aureus* proliferation within *M. sexta* over 72 hours (n = 14).

7.3.5 Virulence varies between natural *S. aureus* isolates

Three paired combinations of isolates were repeated in triplicate (with the exception of D339 and C720) to assess the level of virulence between low and high virulence *S. aureus* strains. The virulence of each strain was determined by i) the number (0-7) of genes present in each strain which were identified by Peacock *et al.*, (2002) as being significantly associated with virulence, and ii) the epidemiological source of the isolate (i.e. whether it derived from asymptomatic carriage or from a case of serious invasive disease (Table 4). Low virulence strains were chosen as those exhibiting 0/7 virulence factors and recovered from asymptomatic carriage. Conversely, high virulence strains were chosen on the basis of exhibiting 7/7 virulence factors and having been recovered from serious invasive disease. The relative fitness index was calculated, and T-test analysis & non-parametric analysis (Mann-Whitney test) performed (Table 5). Strains recovered from invasive disease (containing 7/7 virulence determinants) were significantly more virulent in two of the pairs ($P < 0.01$), with no significant difference in the third pair ($P < 0.05$). The difference in virulence between the two strains selected was more marked when the initial injecting dose was $10^4$ rather than $10^3$ CFU (Fig. 7).

| | 0/7 virulence determinants | 7/7 virulence determinants |
|---|---|---|
| Pair 1 | D339 | C900 |
| Pair 2 | D341 | C640 |
| Pair 3 | D339 | C720 |

**Table 4** Experimental strain combinations.

| Pair Combination | Mann-Whitney analysis | T-test analysis |
|---|---|---|
| D339 $10^3$ & D339 $10^4$ | ($P < 0.01$) | ($P < 0.01$) |
| C900 $10^3$ & C900 $10^4$ | ($P < 0.01$) | ($P < 0.01$) |
| D339 $10^3$ & C900 $10^3$ | ($P < 0.05$) | ($P < 0.05$) |
| D339 $10^4$ & C900 $10^4$ | ($P < 0.01$) | ($P < 0.01$) |
| D341 $10^3$ & D34110$^4$ | ($P > 0.05$) | ($P > 0.05$) |
| C640 $10^3$ & C640 $10^4$ | ($P < 0.01$) | ($P < 0.01$) |
| D341 $10^3$ & C640 $10^3$ | ($P > 0.05$) | ($P > 0.05$) |
| D341 $10^4$ & C640 $10^4$ | ($P < 0.01$) | ($P < 0.01$) |
| D339 $10^3$ & D339 $10^4$ | ($P > 0.05$) | ($P < 0.05$) |
| C720 $10^3$ & C720 $10^4$ | ($P < 0.02$) | ($P < 0.01$) |
| D339 $10^3$ & C720 $10^3$ | ($P > 0.05$) | ($P > 0.05$) |
| D339 $10^4$ & C720 $10^4$ | ($P > 0.05$) | ($P > 0.05$) |

**Table 5** Strains recovered from cases of invasive disease (prefixed C) and exhibiting all 7 virulence genes are significantly more virulent than strains recovered from asymptomatic carriage (prefixed D).

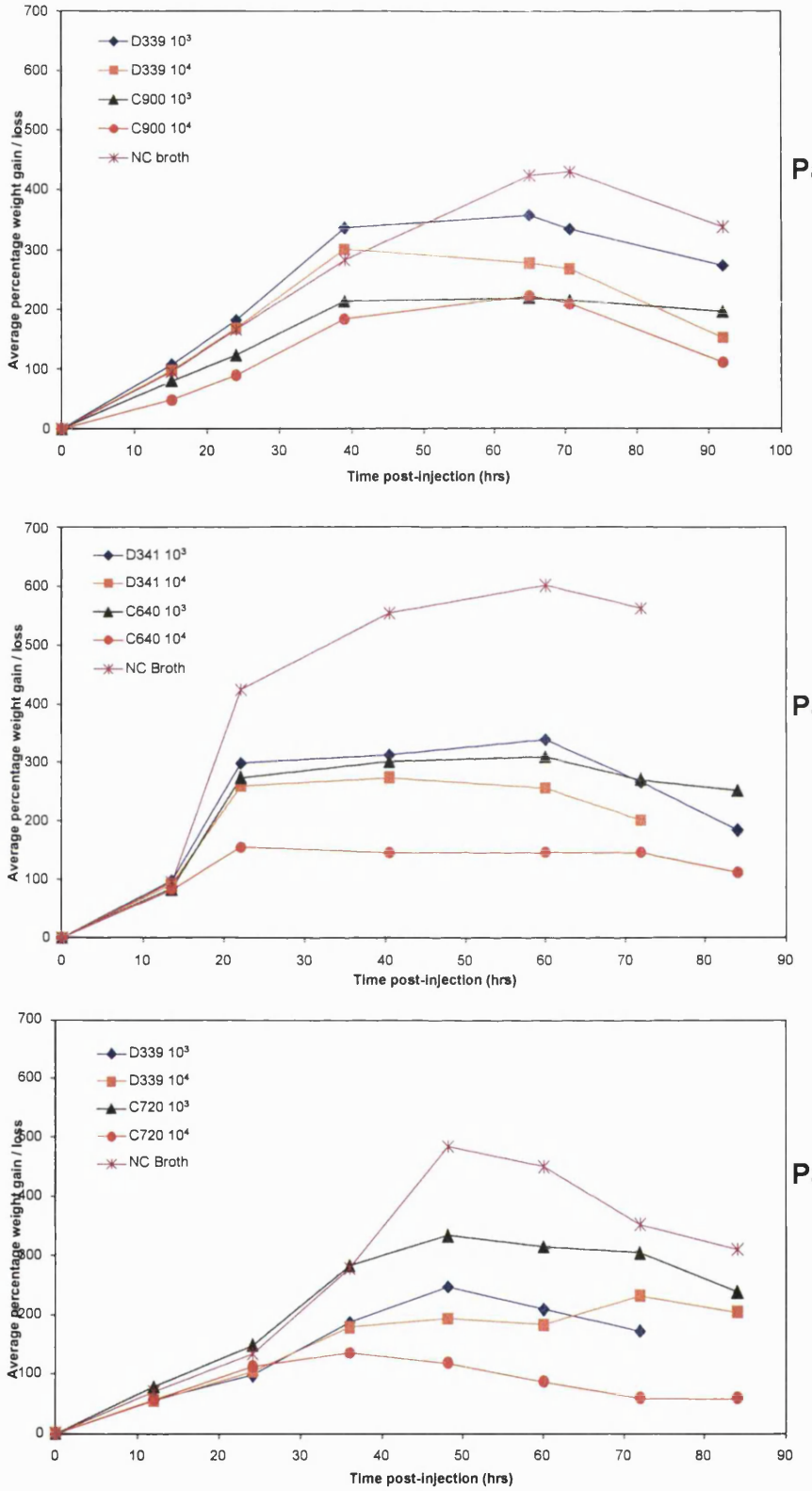



**Figure 7** Significant differences in virulence, as measured by percentage weight gain / loss between "low" virulence and "high" virulence *S. aureus* is observed.

### 7.3.6 Intermediate virulence determinant strains investigated

To examine whether the observed difference in virulence was cumulative, D52, an intermediate virulence strain, was selected (5/7 virulence determinants) to investigate the relative fitness index of "intermediate" strains. *M. sexta* larvae were injected with 10 µl of D52 & H182 (5/7 & 7/7) and average percentage weight gain/loss monitored Figure 8 shows the average percentage weight gain for strain H182 being lower than that of D52 at dose $10^4$ CFU, therefore it can be presumed that this strain is more virulent than D52. It is important to take into account that at the lower injecting dose ($10^3$ CFU) a difference is observed between the two strains in percentage weight gain, however, at $10^4$ CFU this difference is markedly reduced. The data suggest that strains exhibiting an intermediate number of the 7 virulence factors may show intermediate virulence in larvae, however further experimentation is needed to statistically confirm this result.

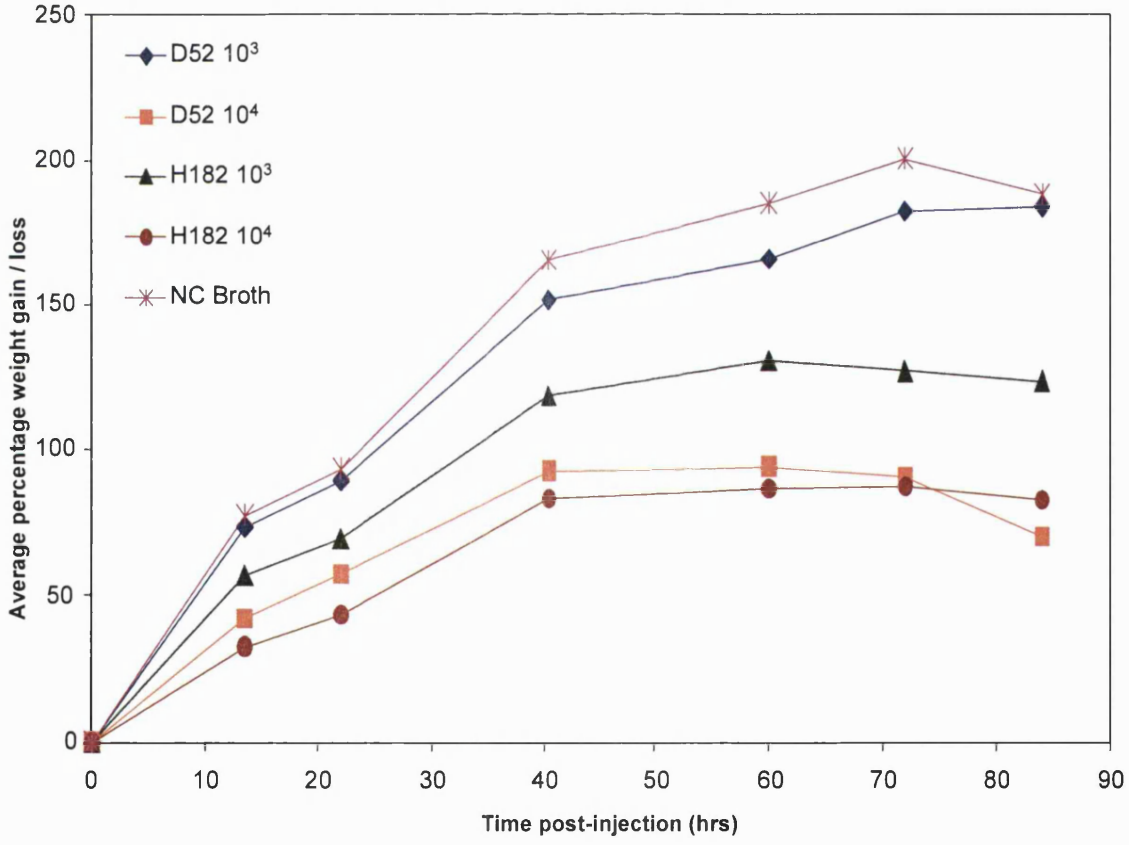



**Figure 8** *S. aureus* strains exhibiting an intermediate number of the 7 virulence factors show intermediate virulence in *M. sexta*.

### 7.3.7 Passage of *S. aureus in vivo*

Passaging of *S. aureus in vivo* was performed to track changes in virulence determinants at the genomic level, and to determine whether the level of virulence increases as the bacteria becomes adapted to the insect host. Initial experimentation, using passaging protocol 1, revealed some larvae cleared the infection before the subsequent passage (high weight gain; data not shown) therefore an injecting dose < $10^3$ CFU was not maintained at each passage. As a result of this only two of the fifteen passaging runs progressed to completion. To take into account the fact that some larvae may clear the *S. aureus* infection, passaging protocol 2 was subsequently used. By injecting larvae with a calculated dose of $10^3$ CFU a continuous set of passaging for 15 (3 x n=5) larvae was easily achieved.

MLST analysis was performed on the two fully passaged isolates from protocol 1 (i.e. each isolate completed 9 continuous passages) to track potential changes in housekeeping loci. Generation of nucleotide sequence data showed no changes in the allelic profile of passaged strains when compared to the ancestral strain. Further experimentation should be performed to track changes in outer surface proteins, toxin genes and the *agr* locus.

### 7.3.8 Agr interference between clinical *S. aureus* strains

A primary aim of this study was to identify genetic factors that correlated with *S. aureus* virulence and growth potential *in vivo*. A variety of factors were fitted to a General linear model (GLM) factor (i.e. presence of specific genes, sequence type, and resistance to different antibiotics) simultaneously, and the only factor that contributed to increased growth potential (fitness) *in vivo* was Agr type ($P < 0.01$). The Agr quorum sensing system regulates the expression of many genes in a cell-density-dependent manner (Otto, 2004; Winzer and Williams, 2001). Recent work has established both the existence of multiple (four) *S. aureus* Agr types (Jarraud *et al.*, 2000; Ji *et al.*, 1997; Lyon and Novick, 2004) and at least 20 others in related species, where interference between different Agr types has been observed (Ji *et al.*, 1997). The marked strain was of Agr type 2, and we found that the fitness of competing strains went in the order of Agr1, Agr2 and Agr3 (Table 6). This suggests either that certain Agr types are more successful at growing in this

environment, or, given the nature of the experiment (i.e. two strains competing within the same insect) Agr interference between the competitor and marked strains is occurring within the insect.

| | | **Relative fitness of competitor strains** | | |
|---|---|---|---|---|
| | | Agr1 | Agr2 | Agr3 |
| **Agr type of marked strain** | Agr1 | 1 | 1.18 ± 0.12 | 1.31 ± 0.17 |
| | Agr2 | 1.25 ± 0.06 | 1 | 0.84 ± 0.06 |
| | Agr3 | 1.12 ± 0.17 | 1.06 ± 0.14 | 1 |

**Table 6** Agr interference occurs between competing *S. aureus* strains during infection. Table showing the mean relative fitness of the different Agr strains when competed against marked strains of each of Agr type 1, 2 and 3 (± the standard error of the mean).

If interference occurred between specific Agr types, we would expect to see the order of fitness of the different Agr types to vary depending on the Agr type they were competing against. Note that the large diversity of genetic backgrounds within this collection should, if anything, obscure any such interactions. By contrast, if Agr type was simply correlated with different absolute competitive abilities, the order of fitness of Agr types should remain the same, regardless of the marked strain. To address this, we repeated the competitions against marked strains of Agr types 1 and 3. We found that relative fitness within the insect was dependent upon the combination of the Agr types of the competing strains, where the ascending order of fitness against marked strains 2 and 3 was 3-2-1, but was 1-2-3 against marker strain 1 (Table 6), suggesting that Agr types interfere with each other. It is however possible that other factors in linkage disequilibrium with the Agr type, not evident through genetic analysis of the strain collection, could also be responsible for the observed interference.

Interestingly, no effect of Agr type was detectable when competitions were performed in BHI broth. That no effect of Agr on competitive ability was observed under these

conditions suggests that factors affected by Agr such as toxicity and adhesiveness only contribute to fitness *in vivo*. This is consistent with the role of Agr activation during infection where we also found that an Agr mutant of *S. aureus* was significantly attenuated in *M. sexta* (strain PC6911 (8325-4 *agr*Δ::*tet*) a generous gift from Prof. Simon Foster). After 24 hours the larvae (n= 20) infected with the wild type strains lost on average 0.49g, whereas the larvae infected with the Agr mutant gained 0.53g (2-sample *t*-test: t= 5.53, $P<0.0001$). After 72 hours 100% of the larvae infected with the wild type strain were dead, whereas only 20% of those infected with the Agr mutant strain were dead ($P<0.001$).

Previous work has looked at the inhibitory affect of supernatant containing AIPs, or purified/synthesized AIPs on Agr activation of *in vitro* populations of *S. aureus* (Jarraud *et al.*, 2000; Ji *et al.*, 1997), or examined how the normal flora of nasal cavities change over time (Kahl *et al.*, 2003; Lina *et al.*, 2003). This study advances these works by demonstrating that Agr interference can occur within a mixed population *in vivo*, and that the competitive ability of a given Agr type depends on what Agr type it is competing against.

## 7.4 DISCUSSION

Insect model systems are becoming increasingly popular alternatives to mammalian model systems for the investigation of virulence mechanisms in human pathogens (Sifri *et al.,* 2003; Mahajan-Miklos *et al.,* 2000). This study aims to develop *Manduca sexta* (the Tobacco Horn Worm), as a potential insect model of infection for the important human pathogen *Staphylococcus aureus.*

Initial *in vivo* studies were aimed to confirm a pilot study by Daniel Bridges and develop the model system. Firstly, it was shown that the injection of sterile TSB did not interfere in *M. sexta* development in any way; each larva progressed to the wandering stage and showed the same weight gain as control larvae that had not been injected. Secondly, preliminary results indicated *S. aureus* kills *M. sexta* with an $LD_{50}$ of $10^4$ CFU (Fig. 6). The initial delay in *M. sexta* death can be attributed to its innate immune response or the time taken for the inoculate to multiply sufficiently for symptoms of infection to occur. Following injection, the integument and gut response to infection and the cellular and humoral defences become activated (Gillespie *et al.,* 1997; Lavine & Strand, 2002).

To confirm colonisation by *S. aureus* is associated with mortality within *M. sexta* death, larvae were injected with live and heat-treated *S. aureus.* Normal development occurred in larvae injected with heat-treated *S. aureus,* indicating that the symptoms of disease in the insects are associated with colonisation by live *S. aureus* cells. This suggestion is also confirmed by plotting growth curves of the inoculate at different time points post infection, and colonisation of *S. aureus* has previously been demonstrated in other non-mammalian hosts (Sifri *et al.,* 2003; Kaito *et al.,* 2002).

Having established that *S. aureus* is able to colonise *M. sexta,* with the onset of disease symptoms, the model was used to investigate i) whether there are any reproducible differences in virulence in the insect host between different naturally occurring strains of *S. aureus*, and ii) to what extent these differences correspond to the known virulence gene profile of the strains and their ability to cause disease in humans. To this end, two pairs of

*S. aureus* strains were compared, in each case multiple experiments were conducted each using 5-10 insect larvae, thus controlling for any differences in host genetics and experimental error. In each case a "carriage" isolate was compared to an isolate recovered from a patient with disease. These strains were also chosen to represent the extremes of virulence gene profile as discussed by Peacock *et al.* having either zero or all seven of the virulence genes these authors showed to be significantly associated with disease onset (Peacock *et al.*, 2002). In both comparisons, significant differences in weight gain are noted post 48-hours after infection, with hosts infected with the "carriage" isolate showing fewer signs of infection. Interestingly, comparisons using the same five pairs of strains were carried out using a nematode host, *Caenorhabditis elegans* by Dr Costi Sifri of Harvard Medical School, and results from these experiments are broadly consistent with those using *M. sexta*, in that the strain C900 and C640 were also shown to be consistently more virulent than D339 than D341 respectively. As the high virulence strain in each pair was recovered from invasive disease ("C" strains), and the low virulence from asymptomatic carriage ("D" strains), these results combined suggest a consistent difference in virulence potential between different *S. aureus* strains when assayed on a remarkable range of unrelated hosts: humans, insects and nematodes. This suggests the involvement of some very ancient and highly conserved proteins or pathways.

It is likely that the observed virulence is due to toxin production, although the genetic basis for pathogenicity in these insect hosts is unclear at present and the broad action of certain adhesins (such as encoded by *sdrE*) cannot be ruled out at this stage. Agr interference experiments demonstrate the role that global regulators, such as the *agr* locus, play in virulence within a mixed population *in vivo*, and that the competitive ability of a given Agr type depends on the Agr type it is competing against (Fleming *et al.*, 2006). The role of the seven virulence factors, shown by Peacock *et al.* to be significantly associated with invasive disease, was challenged recently by a comprehensive microarray analysis on the same strain collection, which failed to confirm that the presence of these genes is significantly associated with disease (Lindsay *et al.*, 2006). As *S .aureus* is known to be a very clonal species (Feil *et al.*, 2003), it is possible that the consistent differences between virulence noted between strains C900 and D339 are due to other factors, acting singly or in

combination, which are linked in the genomes of these strains. In other words, virulence appears to be associated with ST (sequence type), which reflects the clonal background, but not necessarily with specific virulence factors.

Interestingly, the suggestion that the same virulence factors can cause disease symptoms in man and insects is also supported by recent examples of clinical infections of *Photorhabdus* spp. in America and Australia. *Photorhabdus* is a natural pathogen of *M. sexta*, and is commonly vectored by common soil nematodes. Clinical cases of *Photorhabdus* infection in humans appear to have been caused by a distinct sub-species of *Photorhabdus*, which has acquired the ability to infect humans, but has also retained full virulence in insects (Gerrard *et al.*, 2006). The ancient evolutionary role of insect pathogenesis as a source for the emergence of human infection has also been discussed in relation to the important pathogens *Yersinia pestis* and *Bacillus anthracis* (Waterfield *et al.*, 2004).

In addition to providing valuable evidence regarding the potential ability of pathogens to cause clinical symptoms in a wide range of hosts, the *M. sexta* model may also yield important evidence regarding the adaptation of bacteria to a novel *in vivo* environment. During human infection, microorganisms are placed under "stressful" conditions (e.g. alternative pH & temperatures, immune modulators), which may interfere with host colonisation. Within this period they may express/downregulate numerous virulence determinants in order to colonise the surrounding environment (Smith *et al.*, 2006). Passaging experimentation in *M. sexta* aims to mimic these stressful conditions within humans, with each passage representing a new host infection, and it is hoped that genetic changes through sequential passage may be found by sequencing those genes, such as *agr* and genes encoding proteins associated with the cell wall, that may play a pivotal role in the infection process. Initial studies have shown no sequence changes in a sample of housekeeping genes. Further work should aim to demonstrate adaptation in the host by comparing the growth rate of passaged strains within *M. sexta* to their non-passaged ancestors.

# 8. OVERVIEW AND CONCLUDING REMARKS

The assessment of microbial biodiversity and the mechanisms by which it has evolved, is one of the most challenging and fascinating aspects of microbiology. To understand the interaction of microorganisms with their environment or host is fundamental for the development of a proper comprehension of the genetic structure of any bacterial population. This thesis has investigated the natural population structure of two genera, *Wolbachia* spp. and *Vibrio* spp., which illustrate different aspects of bacterial adaptation to eukaryotic hosts, and described an experimental system for assaying natural variation in virulence potential in *Staphylococcus aureus*.

Firstly, this work has illustrated that multi-locus sequence analysis (MLSA) is a robust tool for delineating bacterial species amongst genera, providing greater discriminatory power than other traditional approaches. Both MLSA studies were enhanced by the incorporation of ecological data, allowing more meaningful taxonomic assignments and illustrating the extent of gene flow within and between taxa; this will be discussed further a little later.

Secondly, both *Wolbachia* spp. and *Vibrio* spp. show varying degrees of host adaptation. As an endosymbiont, *Wolbachia* spp. is considered to be under extreme selective pressures to maintain infection in its arthropod host, because it is unculturable outside its host. Consequently, such a vertically transmitted lifestyle might predict co-speciation between host and bacterium. Chapter 4 illustrated host association within sedentary hosts may encourage the emergence of locally adapted clones and hence geographical structuring. At a global scale, frequent horizontal transfer has resulted in a global mosaic genome structure showing little or no congruence between arthropod hosts. Chapters 5 and 6 demonstrate the extent to which *Vibrio* spp., a free-living environmental species, shows geographic structure and how selection and host association contribute to the emergence of locale-specific clones. The data presented are consistent with extreme geographical structure, both on a global level between the UK and NZ populations, and on a more local level (between the two NZ sites), although the project highlights the problem of separating spatial and temporal (seasonal) variation in natural populations.

Both the *Wolbachia* spp. and *Vibrio* spp. populations are characterised by frequent recombination events, testifying to the fundamental importance of this process for bacterial evolution in general, and not just of bacterial pathogens. The rapid expansion of adaptive *Vibrio* clones (i.e. *V. cyclitrophicus* within the UK and *V. pomeroyi* within NZ) also illustrates that recombinogenic populations can be highly dynamic. The use of *ompK* within the *Vibrio* MLSA, presents an ideal marker for measuring host specialisation and interspecies recombination, as this gene is involved in interactions with the host environment, as it encodes a receptor for the broad host-range Vibriophage KVP40, and thus provides evidence concerning the co-evolution of bacteria with phages, complimenting evidence concerning host-association.

The question of how bacteria are able to overcome species barriers and adapt to new hosts is central to the understanding of both the origin of infectious diseases and the emergence of new pathogens. Local adaptation has a direct bearing on the threat posed by *Vibrio* spp. to human health. If local adaptation is minimal, such that it is not possible to predict the geographical origin of an isolate from its genotype, then the implications of high rates global migration would heighten its threat. Conversely, low levels of migration and local ecological adaptation should reduce the risk of novel and potentially dangerous recombinants emerging in the wild.

The hypothesis of Baas-Becking and Beijerinck, who stated *"Everything is everywhere, but the environment selects"* is one that is widely debated by microbiologists today (Beijerinck, 1913). *Wolbachia* spp. and *Vibrio* spp. support this view, however each to differing degrees. This statement is more difficult to apply to *Wolbachia*, as the bacterium is restricted to its host, however within Chapter 4, host congruence is observed within the population (*"Everything is everywhere..."*) with selective sweeps introduced throughout the population by host adaptation according to geographic location (*"...but the environment selects"*). With host incongruence, little if any geographic structure and moderate levels of horizontal transmission observed in Chapter 3, this hypothesis is difficult to interpret at a global level. In Chapters 5 and 6, *Vibrio* spp. fulfil both aspects of this hypothesis. Fifteen *Vibrio* species were identified globally, of which only five were common to both UK and NZ. Species not common to both sites have therefore been subject to "historical" events which have shaped the population in each environment accordingly. A number of isolates, from both UK and

NZ, possess "cosmopolitan" alleles (i.e. they were observed globally), which implies frequent migration from one location to another by passive dispersal, followed by selection to a new environment.

An *in vivo* model of virulence was established in Chapter 7 with pairs of *S. aureus* strains chosen which represented the extremes of virulence gene profile, as discussed by Peacock *et al.*, (2002). In each case a "carriage" isolate was compared to an isolate recovered from a patient with disease. Significant differences in virulence were observed in selected pairs of isolates, in that the strain C900 was also shown to be consistently more virulent than D339 and as was C640 and D341 respectively. These conclusions were corroborated by similar killing experiments using *Caenorhabditis elegans*. Reproducible differences in virulence were however not successful in remaining pairs tested. A recent microarray analysis has subsequently identified the seven loci outlined by Peacock *et al.*, (2002) are not significantly associated with disease, therefore the significant and consistent differences between virulence can be attributed most likely to the sequence type of the strain, which reflects clonal background, but not necessarily specific virulence factors. Agr interference experiments demonstrate the role that global regulators, such as the *agr* locus, play in virulence within a mixed population *in vivo*, and that the competitive ability of a given Agr type depends on the Agr type it is competing against (Fleming *et al.*, 2006).

This experimental work confirms that *S. aureus* is able to colonise and "infect" *M. sexta*, an invertebrate host with which is not known to be associated in nature, and previous work has shown that this species is also able to infect nematodes. In each of the three genera studied in this thesis: *Wolbachia*, *Vibrio* and *Staphylococcus*, it is essentially the lack of specific host association and the ability to interact with a very wide range of hosts, or to rapidly "jump" from one host type to another, that is the most striking and consistent theme. This is consistent with the emerging dogma from bacterial genomics that ecological specialisations can be acquired very quickly through the horizontal transfer of DNA. If this proves to be a general rule, then the role of host associations in dictating bacterial phylogeny and geographic structure might be more limited than previously imagined.

# LITERATURE CITED

**Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L. and Polz, M. F.** (2004). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**(6999), 551-554.

**Alegado R.A., Campbell M.C., Chen W.C., Slutz S.S. and Tan M.** (2003). Characterization of mediators of microbial virulence and innate immunity using the *Caenorhabditis elegans* host-pathogen model. *Cell Microbiol.* **5**(7); 435–444.

**Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. and Lipman, D. J.** (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**(3), 403-410.

**Anderson, C. and Karr, T.** (2001). *Wolbachia*: Evolutionary novelty in a rickettsial bacteria. *BMC Evol. Biol.* **1**(1), 10.

**Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. and Koonin, E. V.** (1998). Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends in Genet.* **14**(11), 442-444.

**Attenborough, D.** (2006). Caves. In *Planet Earth*, (ed. A. Fothergill). London, UK: BBC.

**Baigent, S. J. and McCauley, J. W.** (2003). Influenza type A in humans, mammals and birds: Determinants of virus virulence, host-range and interspecies transmission. *BioEssays* **25**(7), 657-671.

**Baldo, L., Bordenstein, S. R., Wernegreen, J. J. and Werren, J. H.** (2005a). Widespread Recombination Throughout *Wolbachia* Genomes. *Mol. Biol. Evol.* **23**(2), 437 - 449.

**Baldo, L., Lo, N. and Werren, J. H.** (2005b). Mosaic Nature of the *Wolbachia* Surface Protein. *J. Bact.* **187**(15), 5406-5418.

**Bamshad, M., Wooding, S., Salisbury, B. A. and Stephens, J. C.** (2004). Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* **5**(8), 598-609.

**Bandelt, H. J. and Dress, A. W.** (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol Phylogenet. Evol.* **1**(3), 242 - 252.

**Bandi, C., Anderson, T. J. C., Genchi, C. and Blaxter, M. L.** (1998). Phylogeny of *Wolbachia* in filarial nematodes. *Proc. R. Soc. Lond. B* **265**(1413), 2407.

**Banin, E., Israely, T., Kushmaro, A., Loya, Y., Orr, E. and Rosenberg, E.** (2000). Penetration of the Coral-Bleaching Bacterium *Vibrio shiloi* into *Oculina patagonica*. *Appl. Environ. Microbiol.* **66**(7), 3031-3036

**Barbieri, E., Falzano, L., Fiorentini, C., Pianetti, A., Baffone, W., Fabbri, A., Matarrese, P., Casiere, A., Katouli, M., Kuhn, I., Mollby, R., Bruscolini, F., and Donelli, G.** (1999). Occurrence, Diversity, and Pathogenicity of Halophilic *Vibrio* spp. and Non-O1 Vibrio cholerae from Estuarine Waters along the Italian Adriatic Coast. *Appl. Environ. Microbiol.* **65**(6), 2748-2753.

**Bartsch, I.** (2004). Geographical and ecological distribution of marine halacarid genera and species (Acari: Halacaridae). *Exp. Appl. Acarol.* **34**(1-2), 37 - 58.

**Brauns, L., Hudson, M. and Oliver, J.** (1991). Use of the polymerase chain reaction in detection of culturable and nonculturable *Vibrio vulnificus* cells. *Appl. Environ. Microbiol.* **57**(9), 2651-2655.

**Brayton, P., Tamplin, M., Huq, A. and Colwell, R.** (1987). Enumeration of *Vibrio cholerae* O1 in Bangladesh waters by fluorescent-antibody direct viable count. *Appl. Environ. Microbiol.* **53**(12), 2862-2865.

**Beeder, J., Nilsen, R. K., Rosnes, J. T., Torsvik, T. and Lien, T.** (1994). Archaeoglobus fulgidus Isolated from Hot North Sea Oil Field Waters. *Appl. Environ. Microbiol.* **60**(4), 1227 - 1231.

**Beijerinck, M. W.** (1913). De infusies en de ontdekking der bacterien. Amsterdam: Muller.
**Berry, A. J., Ajioka, J. W. and Kreitman, M.** (1991). Lack of Polymorphism on the Drosophila Fourth Chromosome Resulting From Selection. *Genetics* **129**(4), 1111-1117.

**Blindeman, L. and Van Labeke, M. C.** (2003). Control of the two-spotted spider mite (*Tetranychus urticae* Koch) in glasshouse roses. *Commun. Agric. Appl. Biol. Sci.* **68**(4 Pt A), 249 - 264.

**Bordenstein, S. and Rosengaus, R.** (2005). Discovery of a Novel *Wolbachia* Supergroup in Isoptera. *Curr. Microbiol.* **51**(6), 393-398.

**Bordenstein, S. R. and Wernegreen, J. J.** (2004). Bacteriophage Flux in Endosymbionts (*Wolbachia*): Infection Frequency, Lateral Transfer, and Recombination Rates. *Mol. Biol. Evol.* **21**(10), 1981-1991.

**Bouchon, D., Rigaud, T. and Juchault, P.** (1998). Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proc. R. Soc. Lond. B* **265**(1401), 1081-1090.

**Bourtzis, K., Nirgianaki, A., Markakis, G. and Savakis, C.** (1996). *Wolbachia* Infection and Cytoplasmic Incompatibility in *Drosophila* Species. *Genetics* **144**(3), 1063-1073.

**Braig, H. R., Zhou, W., Dobson, S. L. and O'Neill, S. L.** (1998). Cloning and Characterization of a Gene Encoding the Major Surface Protein of the Bacterial Endosymbiont *Wolbachia pipientis. J. Bacteriol.* **180**(9), 2373-2378.

**Brayton, K. A., Kappmeyer, L. S., Herndon, D. R., Dark, M. J., Tibbals, D. L., Palmer, G. H., McGuire, T. C. and Knowles, D. P., Jr.** (2005). Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *PNAS* **102**(3), 844-849.

**Breeuwer, J. A. and Jacobs, G.** (1996). *Wolbachia*: intracellular manipulators of mite reproduction. *Exp Appl Acarol.* **20**(8), 421 - 434.

**Bridges D.** (2002) Attempts To Establish And Characterise An Insect Model For *Staphylococcus aureus* Infection. *Report No. 4/02, University of Bath.*

**Brooks, G. F., Butel, J. S. and Morse, S. A.** (1998). Jawetz, Melnick, & Adelberg's Medical Microbiology. New Jersey, USA: Prentice Hall.

**Brown, T. A.** (2002). Genomes. Oxford, UK: BIOS Scientific Publishers Ltd.

**Buchrieser, C., Gangar, V., Murphree, R., Tamplin, M. and Kaspar, C.** (1995). Multiple *Vibrio vulnificus* strains in oysters as demonstrated by clamped homogeneous electric field gel electrophoresis. *Appl. Environ. Microbiol.* **61**(3), 1163-1168.

**Bullock, G. L.** (1987). Fish Disease Leaflet 77: Vibriosis in Fish, vol. 2006 (ed.: U.S. Fish and Wildlife Service).

**Burnip, G., Charles, J., Shaw, P., Suckling, M., Thomas, W., Tomkins, A., Walker, J. and Wearing, H.** (2006). HortNET, vol. 2006 (ed.: The Horticulture and Food Research Institute of New Zealand Limited).

**Casiraghi, M., Bordenstein, S. R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J. J., Werren, J. H. and Bandi, C.** (2005). Phylogeny of *Wolbachia pipientis* based on *gltA, groEL* and *ftsZ* gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the Wolbachia tree. *Microbiology* **151**(12), 4015-4022.

**Caturegli, P., Asanovich, K. M., Walls, J. J., Bakken, J. S., Madigan, J. E., Popov, V. L. and Dumler, J. S.** (2000). *ankA*: an *Ehrlichia phagocytophila* Group Gene Encoding a Cytoplasmic Protein Antigen with Ankyrin Repeats. *Infect. Immun.* **68**(9), 5277-5283.

**Chan, V. L.** (2003). Bacterial Genomes and Infectious Diseases. *Pediatric Research*(54), 1-7.

**Charlat, S., Hurst, G. D. D. and Mercot, H.** (2003). Evolutionary consequences of *Wolbachia* infections. *Trends in Genet.* **19**(4), 217-223.

**Christensen, H., Bisgaard, M., Frederiksen, W., Mutters, R., Kuhnert, P. and Olsen, J.** (2001). Is characterization of a single isolate sufficient for valid publication of a new genus or species? Proposal to modify Recommendation 30b of the Bacteriological Code (1990 Revision). *Int J Syst Evol Microbiol* **51**(6), 2221-2225.

**Cloern, J., Cole, B., Wong, R. and Alpine, A.** (1985). Temporal dynamics of estuarine phytoplankton: A case study of San Francisco Bay. *Hydrobiologia* **129**153-176.

**Coenye, T., Gevers, D., de Peer, Y. V., Vandamme, P. and Swings, J.** (2005). Towards a prokaryotic taxonomy. *FEMS Microbiology Reviews* **29**(147 - 167).

**Cohan, F. M.** (2002). What are bacterial species? *Annual Review of Microbiology* **56**(1), 457-487.

**Colwell, R. R.** (1996). Global Climate and Infectious Disease: The Cholera Paradigm*. *Science* **274**(5295), 2025-2031.

**Colwell, R. R.** (2000). Viable but nonculturable bacteria: a survival strategy. *Journal of Infection and Chemotherapy* **6**(2), 121-125.

**Combes, C.** (2000). Selective pressure in host-parasite systems. *J. Soc. Biol.* **194**(1), 19 - 23.

**Combes, C. and Theron, A.** (2000). Metazoan parasites and resource heterogeneity: constraints and benefits. *International Journal for Parasitology* **30**(3), 299-304.

**Cooper, J. E. and Feil, E. J.** (2004). Multilocus sequence typing - what is resolved? *Trends in Microbiology* **12**(8), 373-377.

**Cottrell, M. T., Wood, D. N., Yu, L. and Kirchman, D. L.** (2000). Selected Chitinase Genes in Cultured and Uncultured Marine Bacteria in the alpha - and gamma -Subclasses of the Proteobacteria. *Appl. Environ. Microbiol.* **66**(3), 1195-1201.

**Croft, B. A. and Jung, C.** (2001). Phytoseiid dispersal at plant to regional levels: a review with emphasis on management of Neoseiulus fallacis in diverse agroecosystems. *Exp Appl Acarol.* **25**(10-11), 763 - 784.

**Croft, B. A. and Luh, H. K.** (2004). Phytoseiid mites on unsprayed apple trees in Oregon, and other western states (USA): distributions, life-style types and relevance to commercial orchards. *Exp Appl Acarol.* **33**(4), 281 - 326.

**Czarnetzki, A. B. and Tebbe, C. C.** (2004). Detection and phylogenetic analysis of *Wolbachia* in Collembola. *Environmental Microbiology* **6**(1), 35-44.

**Dale, J. W.** (1995). Molecular Genetics of Bacteria. Chichester, uk: John Wiley & Sons.

**Darwin, C.** (1859). On the origin of species. London: Murray.

**Davison, J.** (1999). Genetic Exchange between Bacteria in the Environment. *Plasmid* **42**(2), 73-91.

**Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M. E. and Bouletreau, M.** (2001). Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *PNAS* **98**(11), 6247-6252.

**del Refugio Castaneda Chavez, M., Sedas, V. P., Orrantia Borunda, E. and Reynoso, F. L.** (2005). Influence of water temperature and salinity on seasonal occurrences of *Vibrio cholerae* and enteric bacteria in oyster-producing areas of Veracruz, Mexico. *Marine Pollution Bulletin* **50**(12), 1641-1648.

**DePaola, A., Nordstrom, J. L., Bowers, J. C., Wells, J. G. and Cook, D. W.** (2003). Seasonal Abundance of Total and Pathogenic *Vibrio parahaemolyticus* in Alabama Oysters. *Appl. Environ. Microbiol.* **69**(3), 1521-1526.

**DIVERSITAS.** (2006). DIVERSITAS: an international programme of biodiversity science, vol. 2006.

**Dobson, S. L., Marsland, E. J., Veneti, Z., Bourtzis, K. and O'Neill, S. L.** (2002). Characterization of *Wolbachia* Host Cell Range via the In Vitro Establishment of Infections. *Appl. Environ. Microbiol.* **68**(2), 656-660.

**Doolittle, W. F.** (1999). Lateral genomics. *Trends in Cell Biology* **9**(12), M5-M8.

**Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S.** (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucl. Acids Res.* **19**(14), 4008.

**Drouin, G., Prat, F., Ell, M. and Clarke, G.** (1999). Detecting and characterizing gene conversions between multigene family members. *Mol Biol Evol* **16**(10), 1369-1390.

**Dumler, J., Barbet, A., Bekker, C., Dasch, G., Palmer, G., Ray, S., Rikihisa, Y. and Rurangirwa, F.** (2001). Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six

new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int J Syst Evol Microbiol* **51**(6), 2145-2165.

**Dumontet, S., Krovacek, K., Baloda, S. B., Grottoli, R., Pasquale, V. and Vanucci, S.** (1996). Ecological relationship between *Aeromonas* and *Vibrio* spp. and planktonic copepods in the coastal marine environment in Southern Italy. *Comparative Immunology, Microbiology and Infectious Diseases* **19**(3), 245-254.

**Edgar, R. C.** (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**(5), 1792-1797.

**Egan, E. S., Fogel, M. A. and Waldor, M. K.** (2005). Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Molecular Microbiology* **56**(5), 1129-1138.

**Eisen, J. A.** (2000). Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Current Opinion in Genetics & Development* **10**(6), 606-611.

**Enright, M. C., Day, N. P. J., Davies, C. E., Peacock, S. J. and Spratt, B. G.** (2000). Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of Staphylococcus aureus. *J. Clin. Microbiol.* **38**(3), 1008-1015.

**Esteve-Gassent, M. D., Nielsen, M. E. and Amaro, C.** (2003). The kinetics of antibody production in mucus and serum of European eel (*Anguilla anguilla* L.) after vaccination against *Vibrio vulnificus*: development of a new method for antibody quantification in skin mucus. *Fish & Shellfish Immunology* **15**(1), 51-61.

**Fearnhead, P. and Donnelly, P.** (2001). Estimating Recombination Rates From Population Genetic Data. *Genetics* **159**(3), 1299-1318.

**Feil, E. J.** (2004). Small change: keeping pace with microevolution. *Nat Rev Micro* **2**(6), 483-495.

**Feil, E. J., Cooper, J. E., Grundmann, H., Robinson, D. A., Enright, M., Berendt, T., Peacock, S. J., Smith, J. M., Murphy, M., Spratt, B. G., Moore, Catrin E. and Nicholas P.J.** (2003). How Clonal Is *Staphylococcus aureus*? *J. Bacteriol.* **185**(11), 3307-3316.

**Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. and Spratt, B. G.** (2004). eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *J. Bacteriol.* **186**(5), 1518-1530

**Feil, E. J. and Spratt, B. G.** (2001). Recombination and the population structures of bacterial pathogens. *Annu. Rev. Microbiol.* **55**561 - 590.

**Felsenstein, J.** (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* **17**(6), 368 - 376.

**Fleming, V., Feil, E., Sewell, A. K., Day, N., Buckling, A. and Massey, R. C.** (2006). Agr interference between clinical *Staphylococcus aureus* strains in an insect model of virulence. *J. Bacteriol.* JB.00700-06.

**Ford, T. R., Dillehay, D. L. and Mook, D. M.** (2004). Cutaneous acariasis in the African clawed frog (*Xenopus laevis*). *Comp Med.* **54**(6), 713 - 717.

**Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V., Kumar, S., Posfai, J., Vincze, T., Ingram, J., Moran, L., Lapidus, A., Omelchenko, M., Kyrpides, N., Ghedin, E., Wang, S., Goltsman, E., Joukov, V., Ostrovskaya, O., Tsukerman, K., Mazur, M., Comb, D., Koonin, E.and Slatko, B.** (2005). The *Wolbachia* Genome of *Brugia malayi*: Endosymbiont Evolution within a Human Pathogenic Nematode. *PLoS Biology* **3**(4), e121.

**Fox, G. E., Wisotzkey, J. D. and Jurtshuk, P.** (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol.* **42**(1), 166 - 170.

**Francois, P., Koessler, T., Huyghe, A., Harbarth, S., Bento, M., Lew, D., Etienne, J., Pittet, D. and Schrenzel, J.** (2006). Rapid *Staphylococcus aureus agr* Type Determination by a Novel Multiplex Real-Time Quantitative PCR Assay. *J. Clin. Microbiol.* **44**(5), 1892-1895

**Fu, Y. X. and Li, W. H.** (1993). Statistical Tests of Neutrality of Mutations. *Genetics* **133**(3), 693-709.

**Funchain P., Yeung A., Stweart J., Clendenin W.M. and Miller J.H.** (2001) Amplification of Mutator Cells in a Population as a Result of Horizontal Transfer. *J. Bacteriol.* **183**(12); 3737 – 3741.

**Gerrard, J. G., Joyce, S. A., Clarke, D. J., ffrench-Constant, R. H., Nimmo, G. R., Looke, D. F. M., Feil, E. J., Pearce, L. and Waterfield, N.** (2006). Nematode symbiont for *Photorhabdus asymbiotica*. *Emerg. Infect. Dis.* **12**(10), 1562 - 1564.

**Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., de Peer, Y. V., Vandamme, P., Thompson, F. L. and Swings J.** (2005). Opinion: Re-evaluating prokaryotic species. *Nat Rev Micro* **3**(9), 733-739.

**Gillespie J.P., Kanost M.R. and Trenczek T.** (1997) Biological Mediators of Insect Immunology *Annu. Rev. Entomol.* **42**; 611 – 643.

**Giraudo A.T., Mansilla C., Chan A., Raspanti C. and Nagel R.** (2003) Studies on the Expression of Regulatory Locus *sae* in *Staphylococcus aureus*. *Curr Op Microbiol.* **46**; 246 – 250.

**Gojobori, T., Li, W.-H. and Graur, D.** (1982). Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.* **18**360 - 369.

**Goldsby R. A., Kindt T. J. and Osborne B. A.** (2000) Overview of the immune System. In "Kuby Immunology" 4th edn (Folchetti N., Tannenaum J., Halder G.L. & Kudlacik A. eds) **35**; 3 – 27 New York, WH Freeman & Company.

**Gonzalez-Escalona, N., Fey, A., Hofle, M. G., Espejo, R. T. and A. Guzman, C.** (2006). Quantitative reverse transcription polymerase chain reaction analysis of *Vibrio cholerae* cells entering the viable but non-culturable state and starvation in response to cold shock. *Environmental Microbiology* **8**(4), 658-666.

**Gophna, U., Charlebois, R. L. and Doolittle, W. F.** (2004). Have archaeal genes contributed to bacterial virulence? *Trends in Microbiology* **12**(5), 213-219.

**Griffin, D., Kellogg, C., Garrison, V. and Shinn, E.** (2002). The global transport of dust. *Am Sci* **90**228 - 235.

**Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C. and Gelbart, W. M.** (2000). An introduction to genetic analysis, vol. 2006 (ed.: W. H. Freeman and Company).

**Groot, T. V. M. and Breeuwer, J. A. J.** (2006). *Cardinium* symbionts induce haploid thelytoky in most clones of three closely related *Brevipalpus* species. *Exp Appl Acarol.* **39**(3-4), 257 - 271.

**Gupta, S. and Maiden, M. C. J.** (2001). Exploring the evolution of diversity in pathogen populations. *Trends in Microbiology* **9**(4), 181 - 185.

**Guttman, D. S. and Dykhuizen, D. E.** (1994). Detecting Selective Sweeps in Naturally Occurring *Escherichia coli*. *Genetics* **138**(4), 993-1003.

**Hahn, B. H., Shaw, G. M., De Cock, K. M. and Sharp, P. M.** (2000). AIDS as a Zoonosis: Scientific and Public Health Implications. *Science* **287**(5453), 607-614.

**Hall, B. G.** (2004). Phylogenetic Trees Made Easy: A how-to manual. Sunderland, USA: Sinauer Associates.

**Hall, T. A.** (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.*(41), 95 - 98.

**Hallin, P. F. and D.W., U.** (2004). CBS Genome Atlas Database: a dynamic storage for bioinformatic results and sequence data. *Bioinformatics* **20**(18), 3682–3686.

**Hanage, W. P., Fraser, C. and Spratt, B. G.** (2005). Fuzzy species among recombinogenic bacteria. *BMC Biology* **3**(6).

**Hasegawa, M., Kishino, H. and Yano, T.** (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* **22**(2), 160 - 174.

**Hastings, P. J. and Rosenberg, S. M.** (2002). In pursuit of a molecular mechanism for adaptive gene amplification. *DNA Repair* **1**(2), 111-123.

**Heath, B. D.** (2000). The Role of Parasitoids in the Horizontal Transmission of *Wolbachia*. In *Faculty of Science*, (ed., pp. 10 - 16. Dundee: University of Dundee).

**Heath, B. D., Butcher, R. D. J., Whitfield, W. G. F. and Hubbard, S. F.** (1999). Horizontal transfer of *Wolbachia* between phylogenetically distat insect species by a naturally occurring mechanism. *Current Biology*(9), 313 - 316.

**Heidelberg, J., Eisen, J., Nelson, W., Clayton, R., Gwinn, M., Dodson, R., Haft, D., Hertig, M.** (1936). The rickettsia, *Wolbachia pipientis* (gen. et sp. n.) and associated inclusions of the mosquito *Culex pipiens*. *Parasitology*(28), 435 - 486.

**Heidelberg, J., Eisen, J., Nelson, W., Clayton, R., Gwinn, M., Dodson, R., Haft, D., Hickey, E., Peterson, J., Umayam, L., Gill, S., Nelson, K., Read, T., Tettelin, H., Richardson, D., Ermolaeva, M., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R., Nierman, W., White, O., Salzberg, S., O'Smith, H., Colwell, R., Mekalanos, J., Venter, C. and Fraser, C.** (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**477-483.

**Heidelberg, J. F., Heidelberg, K. B. and Colwell, R. R.** (2002). Bacteria of the γ-Subclass Proteobacteria Associated with Zooplankton in Chesapeake Bay. *Appl. Environ. Microbiol.* **68**(11), 5498-5507.

**Hertig, M. and Wolbach, S. B.** (1924). Studies on rickettsia-like microorganisms in insects. *J. Med. Entomol.*(44), 329 - 374.

**Heukelbach, J. and Feldmeier, H.** (2006). Scabies. *The Lancet* **3671**767 - 1774.

**Hilger, H. and Humer, M.** (2003). Biotic landfill cover treatments for mitigating methane emissions. *Environ. Monit. Assess.* **84**(1-2), 71 - 84.

**Hiroki, M., Tagami, Y., Miura, K. and Kato, Y.** (2004). Multiple infection with *Wolbachia* inducing different reproductive manipulations in the butterfly *Eurema hecabe*. *Proc. R. Soc. Lond. B* **271**(1549), 1751-1755.

**Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., Jr. and Ezekowitz, R. A. n. B.** (1999). Phylogenetic Perspectives in Innate Immunity. *Science* **284**(5418), 1313-1318.

**Hooper, S. D. and Berg, O.** (2003). Duplication is more common among laterally transferred genes than among indigenous genes. *Genome Biol.* **4**(8), R48.

**Horst, J.-P., Wu, T.-h. and Marinus, M. G.** (1999). *Escherichia coli* mutator genes. *Trends in Microbiology* **7**(1), 29-36.

**Hountondji, F. C. C., Sabelis, M. W., Hanna, R. and Janssen, A.** (2005). Herbivore-induced Plant Volatiles Trigger Sporulation in Entomopathogenic Fungi: The Case of Neozygites tanajoae Infecting the Cassava Green Mite. *Journal of Chemical Ecology* **31**(5), 1003 - 10021.

**Hryniewicz-Jankowska, A., Czogalla, A., Bok, E. and Sikorsk, A. F.** (2002). Ankyrins, multifunctional proteins involved in many cellular pathways. *Folia Histochem Cytobiol.* **40**(3), 239 - 249.

**Hudson, R. R.** (1987). Estimating the recombination parameter of a finite population model without selection. *Genet Res.* **50**(3), 245 - 250.

**Hudson, R. R.** (2001). Two-Locus Sampling Distributions and Their Application. *Genetics* **159**(4), 1805-1817.

**Hudson, R. R. and Kaplan, N. L.** (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**(1), 147-164.

**Huelsenbeck, J. P., Larget, B., Miller, R. E. and Ronquist, F.** (2002). Potential Applications and Pitfalls of Bayesian Inference of Phylogeny. *Syst Biol.* **51**(5), 673 - 688.

**Huelsenbeck, J. P. and Ronquist, F.** (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8), 754-755.

**Huelsenbeck, J. P., Ronquist, F., Nielsen, R. and Bollback, J. P.** (2001). Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science* **294**(5550), 2310-2314.

**Hugenholtz, P. and Pace, N. R.** (1996). Identifying microbial diversity in the natural environment: A molecular phylogenetic approach. *Trends in Biotechnology* **14**(6), 190-197.

**Hurst, G. D. D. and Jiggins, F. M.** (2000). Male-Killing Bacteria in Insects: Mechanisms, Incidence, and Implications. *Emerg. Infect. Dis.*(6), 328 - 336.

**Hurst, G. and Jiggins, F.** (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences* **272**(1572), 1525-1534.

**Huson, D.** (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**(1), 68-73.

**Huson, D. H. and Bryant, D.** (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Mol Biol Evol* **23**(2), 254-267.

**Iturbe-Ormaetxe, I., Burke, G. R., Riegler, M. and O'Neill, S. L.** (2005). Distribution, Expression, and Motif Variability of Ankyrin Domain Genes in *Wolbachia pipientis*. *J. Bacteriol.* **187**(15), 5136-5145.

**Jain, R., Rivera, M. C. and Lake, J. A.** (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *PNAS* **96**(7), 3801-3806.

**Jain, R., Rivera, M. C., Moore, J. E. and Lake, J. A.** (2002). Horizontal Gene Transfer in Microbial Genome Evolution. *Theoretical Population Biology* **61**(4), 489-495.

**Jamnongluk, W., Kittayapong, P., Baimai, V. and O'Neill, S. L.** (2002). *Wolbachia* Infections of Tephritid Fruit Flies: Molecular Evidence for Five Distinct Strains in a Single Host Species. *Current Microbiology* **45**(4), 255-260.

**Jander G., Rahme L.G. and Ausubel F.M.** (2000) Positive correlstion between Virulence of *Pseudomonas aeruginosa* Mutants in Mice and Insects *J. Bacteriol.* **182**(13); 3843 – 3845.

**Jarraud, S., Lyon, G. J., Figueiredo, A. M. S., Gerard, L., Vandenesch, F., Etienne, J., Muir, T. W. and Novick, R. P.** (2000). Exfoliatin-Producing Strains Define a Fourth *agr* Specificity Group in *Staphylococcus aureus*. *J. Bacteriol.* **182**(22), 6517-6522.

**Jarraud, S., Mougel, C., Thioulouse, J., Lina, G., Meugnier, H., Forey, F., Nesme, X., Etienne, J. and Vandenesch, F.** (2002). Relationships between *Staphylococcus aureus* Genetic Background, Virulence Factors, *agr* Groups (Alleles), and Human Disease. *Infect. Immun.* **70**(2), 631-641.

**Jermyn, W. S. and Boyd, E. F.** (2005). Molecular evolution of *Vibrio* pathogenicity island-2 (VPI-2): mosaic structure among *Vibrio cholerae* and *Vibrio mimicus* natural isolates. *Microbiology* **151**(1), 311-322.

**Jeyaprakash, A. and Hoy, M. A.** (2000). Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Molecular Biology* **9**(4), 393-405.

**Ji, G., Beavis, R. and Novick, R. P.** (1997). Bacterial Interference Caused by Autoinducing Peptide Variants. *Science* **276**(5321), 2027-2030.

**Jiang, S. C. and Fu, W.** (2001). Seasonal Abundance and Distribution of *Vibrio cholerae* in Coastal Waters Quantified by a 16S-23S Intergenic Spacer Probe. *Microbial Ecology* **42**(4), 540-548.

**Jiang, S. C., Louis, V., Choopun, N., Sharma, A., Huq, A. and Colwell, R. R.** (2000b). Genetic Diversity of *Vibrio cholerae* in Chesapeake Bay Determined by Amplified Fragment Length Polymorphism Fingerprinting. *Appl. Environ. Microbiol.* **66**(1), 140-147.

**Jiang, S. C., Matte, M., Matte, G., Huq, A. and Colwell, R. R.** (2000a). Genetic Diversity of Clinical and Environmental Isolates of *Vibrio cholerae* Determined by Amplified Fragment Length Polymorphism Fingerprinting. *Appl. Environ. Microbiol.* **66**(1), 148-153.

**Jiang, S. C. and Paul, J. H.** (1998). Gene Transfer by Transduction in the Marine Environment. *Appl. Environ. Microbiol.* **64**(8), 2780-2787.

**Jiggins, F. M.** (2002). The Rate of Recombination in *Wolbachia* Bacteria. *Mol Biol Evol* **19**(9), 1640-1643.

**Jiggins, F. M., Schulenburg, J. H. G. v. d., Hurst, G. D. D. and Majerus, M. E. N.** (2001). Recombination confounds interpretations of *Wolbachia* evolution. *Proc. R. Soc. Lond. B* **268**(1474), 1423-1427.

**Johnson, F. H. and Shunk, I. V.** (1936). An interesting new species of luminous bacteria. *J. Bacteriol.* **31**585-593.

**Jolley, K. A., Feil, E. J., Chan, M.-S. and Maiden, M. C. J.** (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**(12), 1230-1231.

**Jolley, K. A., Wilson, D. J., Kriz, P., Mcvean, G. and Maiden, M. C. J.** (2005). The Influence of Mutation, Recombination, Population History, and Selection on Patterns of Genetic Diversity in *Neisseria meningitidis*. *Mol Biol Evol* **22**(3), 562-569.

**Jordan, I. K., Rogozin, I. B., Wolf, Y. I. and Koonin, E. V.** (2002). Microevolutionary Genomics of Bacteria. *Theoretical Population Biology* **61**(4), 435-447.

**Jukes, T. H. and Cantor, C. R.** (1969). Evolution of protein molecules. In *Mammalian protein metabolism.*, (ed. H. N. Munro), pp. Pp 21 - 132. New York: Academic Press.

**Kageyama, D., Nishimura, G., Hoshizaki, S. and Ishikawa, Y.** (2002). Feminizing *Wolbachia* in an insect, *Ostrinia furnacalis* (Lepidoptera: Crambidae). *Heredity* **88**(6), 444 - 449.

**Kahl, B. C., Becker, K., Friedrich, A. W., Clasen, J., Sinha, B., von Eiff, C. and Peters, G.** (2003). *agr*-Dependent Bacterial Interference Has No Impact on Long-Term Colonization of *Staphylococcus aureus* during Persistent Airway Infection of Cystic Fibrosis Patients. *J. Clin. Microbiol.* **41**(11), 5199-5201.

**Kaito C., Akimitsu N., Wantabe H. and Sekimizu K.** (2002) Silkworm larvae as an animal model of bacterial infection pathogenic to humans. *Microb Pathog. 32; 183 – 190.*

**Kampelmacher, E. H., van Noorle Jansen, L. M., Mossel, D. A. and Groen, F. J.** (1972). A survey of the occurrence of *Vibrio parahaemolyticus* and *V. alginolyticus* on mussels and oysters and in estuarine waters in the Netherlands. *J. Appl. Bacteriol.* **25**(3), 431-438.

**Kane, E. and Ochoa, R.** (2006). Systematic Entomology Laboratory, vol. 2006 (ed.: United States Department of Agriculture.

**Kaneko, T. and Colwell, R. R.** (1974). Distribution of Vibrio parahaemolyticus and related organisms in the Atlantic Ocean off South Carolina and Georgia. *Applied Microbiology* **28**(6), 1009-1017.

**Kils-Hutten, L., Cheynier, R., Wain-Hobson, S. and Meyerhans, A.** (2001). Phylogenetic reconstruction of intrapatient evolution of human immunodeficiency virus type 1: predominance of drift and purifying selection. *J Gen Virol* **82**(7), 1621-1627.

**Kim, D. S. and Lee, J. H.** (2003). Oviposition model of overwintered adult *Tetranychus urticae* (Acari: Tetranychidae) and mite phenology on the ground cover in apple orchards. *Exp Appl Acarol.* **31**(3-4), 191 - 208.

**Kimbrell D.A. and Beutler B.** (2001) The Evolution and Genetics of Innate Immunity *Nat Rev Genet. 2(4); 256 – 267.*

**Kimura, M.** (1968). Evolutionary rate at the molecular level. *Nature* **217**(129), 624 - 626.

**Kimura, M.** (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* **16**(2), 111 - 120.

**Kimura, M. and Crow, J. F.** (1964). The number of allele that can be contained in a finite population. *Genetics* **49**(4), 725-738.

**Kiser K.B., Cantey-Kiser J.M. and Lee J.C.** (1999) Development and Characterization of a *Staphylococcus aureus* Nasal Colonization Model in Mice. *Infect Immun. 67(10); 5001 – 5006.*

**Kluytmans J., van Belkum A. and Verbrugh H.** (1997) Nasal Carriage of *Staphylococcus aureus*: Epidemiology, Underlying Mechanisms, and Associated Risks. *Clin Microbiol Rev. 10(3); 505 – 520.*

**Kotetishvili, M., Stine, O. C., Chen, Y., Kreger, A., Sulakvelidze, A., Sozhamannan, S. and Morris, J. G., Jr.** (2003). Multilocus Sequence Typing Has Better Discriminatory Ability for Typing *Vibrio cholerae* than Does Pulsed-Field Gel Electrophoresis and Provides a Measure of Phylogenetic Relatedness. *J. Clin. Microbiol.* **41**(5), 2191-2196.

**Kothary, M. H. and Kreger, A. S.** (1985). Purification and characterization of an extracellular cytolysin produced by *Vibrio damsela. Infect. Immun.* **49**(1), 25-31.

**Krantz, G. E., Colwell, R. R. and Lovelace, E.** (1969). *Vibrio parahaemolyticus* from the blue crab Callinectes sapidus in Chesapeake Bay. *Science* **164**(885), 1286-1287.

**Kuberski, T., Flood, T. and Tera, T.** (1979). Cholera in the Gilbert Island. I. Epidemiological features. *Am J Trop Med Hyg.* **28**(4), 677 - 684.

**Kumar, S., Tamura, K. and Nei, M.** (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* **5**(2), 150 - 163.

**Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K.-i. and Nagai, Y.** (2001). Whole genome sequencing of meticillin-resistant *Staphylococcus aureus. The Lancet* **357**(9264), 1225-1240.

**Kurz C.L. and Ewbank J.J.** (2003) *Caenorhabditis Elegans*: An Emerging Genetic Model For The Study Of Innate Immunity. *Nat Rev Genet. 4(5);380 - 390.*

**Kutschera, U. and Niklas, K. J.** (2004). The modern theory of biological evolution: an expanded synthesis. *Naturwissenschaften* **91**(6), 255 - 276.

**Lan, R. and Reeves, P. R.** (2001). When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends in Microbiology* **9**(9), 419 - 424.

**Lanave, C., Preparata, G., Saccone, C. and Serio, G.** (1984). A new method for calculating evolutionary substitution retes. *J Mol Evol.* **10**(1), 86 - 93.

**Larsen, M. H., Blackburn, N., Larsen, J. L. and Olsen, J. E.** (2004). Influences of temperature, salinity and starvation on the motility and chemotactic response of *Vibrio anguillarum*. *Microbiology* **150**(5), 1283-1290.

**Lee, K. and Ruby, E. G.** (1994). Effect of the Squid Host on the Abundance and Distribution of Symbiotic *Vibrio fischeri* in Nature. *Appl. Environ. Microbiol.* **601**565-1571.

**Lehrer, R. I. and Ganz, T.** (1999). Antimicrobial peptides in mammalian and insect host defence. *Current Opinion in Immunology* **11**(1), 23-27.

**Lenski, R. E., Rose, M. R., Simpson, S. C. and Tadler, S. C.** (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**1315 -1341.

**Li, M., Shimada, T., Morris, J. G., Jr., Sulakvelidze, A. and Sozhamannan, S.** (2002). Evidence for the Emergence of Non-O1 and Non-O139 *Vibrio cholerae* Strains with Pathogenic Potential by Exchange of O-Antigen Biosynthesis Regions. *Infect. Immun.* **70**(5), 2441-2453.

**Lavine M.D. and Strand M.R.** (2002) Insect hemocytes and their role in immunity. *Insect Biochem. Mol. Biol.***32**; 1295 – 1309.

**Lina, G., Boutite, F., Tristan, A., Bes, M., Etienne, J. and Vandenesch, F.** (2003). Bacterial Competition for Human Nasal Cavity Colonization: Role of Staphylococcal *agr* Alleles. *Appl. Environ. Microbiol.* **69**(1), 18-23.

**Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., Husain, S. E., Butcher, P. D. and Hinds, J.** (2006). Microarrays Reveal that Each of the Ten Dominant Lineages of *Staphylococcus aureus* Has a Unique Combination of Surface-Associated and Regulatory Genes. *J. Bacteriol.* **188**(2), 669-676.

**Lloyd, A. T. and Sharp, P. M.** (1993). Evolution of the *recA* gene and the molecular phylogeny of bacteria. *J Mol Evol.*(4), 399-407.

**Lo, N., Casiraghi, M., Salati, E., Bazzocchi, C. and Bandi, C.** (2002). How Many *Wolbachia* Supergroups Exist? *Mol Biol Evol* **19**(3), 341-346.

**Long, R. A. and Azam, F.** (2001). Antagonistic Interactions among Marine Pelagic Bacteria. *Appl. Environ. Microbiol.* **67**(11), 4975-4983.

**Louis, C. and Nigro, L.** (1989). Ultrastructural evidence of *Wolbachia* rickettsiales in *Drosophila simulans* and their relationships with unidirectional cross-incompatibility. *J. Invert. Pathol.* **54**(1), 39 - 44.

**Lyon, G. J. and Novick, R. P.** (2004). Peptide signaling in *Staphylococcus aureus* and other Gram-positive bacteria. *Peptides* **25**(9), 1389-1403.

**Lynch, M. and Conery, J. S.** (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **290**(5494), 1151-1155.

**Mader, S. S.** (1996). Biology. Dubuque, USA: Wm. C. Brown.

**Maeda, T., Takabayashi, J., Yano, S. and Takafuji, A.** (2000). Effects of light on the tritrophic interaction between kidney bean plants, two-spotted spider mites and predatory mites, *Amblyseius womersleyi* (Acari: Phytoseiidae). *Exp Appl Acarol.* **24**(5-6), 415 - 425.

**Mahajan-Miklos S., Rahme L.G. and Ausubel F.M.** (2000) Elucidating the molecular mechanisms of bacterial virulence using the non-mammalian hosts. *Mol Microbiol.* **37**(5); 981-988.

**Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I.M., Achtman, M. and Spratt, B.G.** (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *PNAS* **95**(6), 3140-3145.

**Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M. and Yamashita, A.** (2003). Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *The Lancet* **361**(9359), 743-749.

**Malloch, G. and Fenton, B.** (2005). Super-infections of *Wolbachia* in byturid beetles and evidence for genetic transfer between A and B super-groups of *Wolbachia*. *Molecular Ecology* **14**(2), 627-637.

**Marshall, D. and Convey, P.** (2004). Latitudinal variation in habitat specificity of ameronothrid mites (Oribatida). *Exp Appl Acarol.* **34**(1-2), 21 - 35.

**Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P.J., Naeem, S., Ovreas, L., Reysenbach, A., Smith, V.H. and Staley, J.T.** (2006). Microbial biogeography: putting microorganisms on the map. **4**(2), 102-112.

**Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K. and Fasano, A.** (1998). "Black holes" and bacterial pathogenicity: A large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *PNAS* **95**(7), 3943-3948.

**Maynard Smith, J. and Smith, N.** (1998). Detecting recombination from gene trees. *Mol Biol Evol* **15**(5), 590-599.

**May, R.** (2002). Biological Diversity in a Crowded World: Past, Present and Likely Future, vol. 2006 (ed. University of Oxford).

**Mayr, E.** (1942). Systematics and the Origin of Species. New York, USA: Columbia University Press.

**McCann, J., Stabb, E. V., Millikan, D. S. and Ruby, E. G.** (2003). Population Dynamics of *Vibrio fischeri* during Infection of *Euprymna scolopes*. *Appl. Environ. Microbiol.* **69**(10), 5928-5934.

**McCarthy, S. A. and Khambaty, F. M.** (1994). International dissemination of epidemic *Vibrio cholerae* by cargo ship ballast and other nonpotable waters. *Appl. Environ. Microbiol.* **60**(7), 2597-2601.

**McVean, G., Awadalla, P. and Fearnhead, P.** (2002). A Coalescent-Based Method for Detecting and Estimating Recombination From Gene Sequences *Genetics* **160**(3), 1231-1241.

**Messer, W.** (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiology Reviews* **26**(4), 355-374.

**Min, K.-T. and Benzer, S.** (1997). Wolbachia, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *PNAS* **94**(20), 10792-10796.

**Mira, A., Ochman, H. and Moran, N. A.** (2001). Deletional bias and the evolution of bacterial genomes. *Trends in Genetics* **17**(10), 589-596.

**Moran, N. A.** (2002). Microbial Minimalism: Genome Reduction in Bacterial Pathogens. *Cell* **108**(5), 583-586.

**Moran, N. A.** (2003). Tracing the evolution of gene loss in obligate bacterial symbionts. *Current Opinion in Microbiology* **6**(5), 512-518.

**Moran, N. and Mira, A.** (2001). The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biology* **2**(12), research0054.1 - research0054.12.

**Morris, C. E., Bardin, M., Berge, O., Frey-Klett, P., Fromin, N., Girardin, H., Guinebretiere, M.-H., Lebaron, P., Thiery, J. M. and Troussellier, M.** (2002). Microbial Biodiversity: Approaches to Experimental Design and Hypothesis Testing in Primary Scientific Literature from 1975 to 1999. *Microbiol. Mol. Biol. Rev.* **66**(4), 592-616.

**Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. and Peng, Z.-y.** (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* **13**(6), 1435-1448.

**Mouton, L., Dedeine, F., Henri, H., Bouletreau, M., Profizi, N. and Vavre, F.** (2004). Virulence, Multiple Infections and Regulation of Symbiotic Population in the *Wolbachia-Asobara tabida* Symbiosis. *Genetics* **168**(1), 181-189.

**Mushegian A. and Medzhitov R.** (2001) Evolutionary perspective on innate immune recognition. *J. Cell Biol.* **155**(5); 705-710.

**Myers, S. R. and Griffiths, R. C.** (2003). Bounds on the Minimum Number of Recombination Events in a Sample History *Genetics* **163**(1), 375-394.

**Myhr, E., Larsen, J. L., Lillehaug, A., Gudding, R., Heum, M. and Håstein, T.** (1991). Characterization of *Vibrio anguillarum* and closely related species isolated from farmed fish in Norway. *Appl. Environ. Microbiol.* **57**(9), 2750-2757.

**Mylonakis, E. and Aballay, A.** (2005). Worms and Flies as Genetically Tractable Animal Models To Study Host-Pathogen Interactions. *Infect. Immun.* **73**(7), 3833-3841.

**Nair, S. and Simidu, U.** (1987). Distribution and significance of heterotrophic marine bacteria with antibacterial activity. *Appl. Environ. Microbiol.* **53**(12), 2957-2962.

**Nei, M.** (2005). Selectionism and Neutralism in Molecular Evolution. *Mol Biol Evol* **22**(12), 2318-2342.

**Nei, M. and Kumar, S.** (2000). Molecular Evolution and Phylogenetics. New York, USA: Oxford University Press.

**Northup, D. E., Barns, S. M., Yu, L. E., Spilde, M. N., Schelble, R. T., Dano, K. E., Crossey, L. J., Connolly, C. A., Boston, P. J., Natvig, D. O. and Dahm C.N.** (2003). Diverse microbial communities inhabiting ferromanganese deposits in Lechuguilla and Spider Caves. *Environmental Microbiology* **5**(11), 1071-1086.

**Norton, R. A., Bonamo, P. M., Grierson, J. D. and Shear, W. A.** (1988). Oribatid mite fossils from a terrestrial Devonian deposit near Gilboa, New York. *Journal of Paleontology* **62**259 - 268.

**Notley-McRobb L., Pinto R., Seeto S. and Ferenci T.** (2002) Regulation of *mutY* and Nature Mutations in *Escherichia coli* Populations under Nutritional Limitation. *J. Bacteriol.* **184**(3); 739 – 745.

**Novick R.P.** (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol Microbiol.* **48**(6); 1429 – 1449.

**Novick R.P. and Muir T.W.** (1999) Virulence gene regulation by peptides in staphylococci and other Gram-positive bacteria. *Curr. Op. Microbiol.* **2**; 40 – 45.

**Novick, R. P., Ross, H. F., Projan, S. J., Kornblum, J., Kreiswirth, B. and Moghazeh, S.** (1993). Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO reports* **12**(10), 3967 - 3975.

**Nurminsky, D. I.** (2001). Genes in sweeping competition. *Cell Mol. Life Sci.* **58**(1), 125 - 134.

**Ochman, H.** (2001). Lateral and oblique gene transfer. *Current Opinion in Genetics & Development* **11**(6), 616-619.

**Ochman, H., Lawrence, J. G. and Groisman, E. A.** (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**299-304.

**Ochoa, G. and Jaffe, K.** (1999). On Sex, Mate Selection and the Red Queen. *Journal of Theoretical Biology* **199**(1), 1-9.

**O'Neill, S., Giordano, R., Colbert, A., Karr, T. and Robertson, H.** (1992a). 16S rRNA Phylogenetic Analysis of the Bacterial Endosymbionts Associated with Cytoplasmic Incompatibility in Insects. *PNAS* **89**(7), 2699-2702.

**Otto, M.** (2004). Quorum-sensing control in *Staphylococci* - a target for antimicrobial drug therapy? *FEMS Microbiology Letters* **241**(2), 135-141.

**Page, R. D. M. and Holmes, E.** (2000). Molecular Evolution: A Phylogenetic Approach. Oxford, UK: Blackwell Science Ltd.

**Panaram, K. and Marshall, J.** (2006). F supergroup *Wolbachia* in bush crickets: what do patterns of sequence variation reveal about this supergroup and horizontal transfer between nematodes and arthropods? *Genetica.*

**Park, Y. L. and Lee, J. H.** (2005). Impact of twospotted spider mite (Acari: Tetranychidae) on growth and productivity of glasshouse cucumbers. *J Econ Entemol.* **98**(2), 457 - 463.

**Peacock S.J., de Silva S. and Lowy F.D.** (2001) What determines nasal carriage of *Staphylococcus aureus* ? *Trends Microbiol.* **9**(12); 605-610.

**Peacock S.J., Moore C.E., Justice A., Kantzanou M., Story L., Mackie K., O'Neill G. and Day N.P.J.** (2002) Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus. Infect Immun.* **70**(9); 4987-4996.

**Perez-Losada, M., Browne, E. B., Madsen, A., Wirth, T., Viscidi, R. P. and Crandall, K. A.** (2006). Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infection, Genetics and Evolution* **6**(2), 97-112.

**Pond, S. L. K. and Frost, S. D. W.** (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**(10), 2531-2533.

**Prescott, L. M., Harley, J. P. and Klein, D. A.** (1996). Microbiology. Dubuque, USA: Wm. C. Brown Publishers.

**Prischmann, D. A., Croft, B. A. and Luh, H. K.** (2002). Biological control of spider mites on grape by phytoseiid mites (Acari: Tetranychidae, Phytoseiidae): emphasis on regional aspects. *J Econ Entomol.* **95**(2), 340 - 347.

**Proctor, H. and Owens, I.** (2000). Mites and birds: diversity, parasitism and coevolution. *Trends in Ecology & Evolution* **15**(9), 358-364.

**Provvedi, R., Maggi, T., Oggioni, M. R., Manganelli, R. and Pozzi, G.** (2005). Selection and characterization of a promoter for expression of single-copy recombinant genes in Gram-positive bacteria. *BMC Biotechnol.* **5**(3), 1 - 12.

**Papke, R. T. and Ward, D. M.** (2004). The importance of physical isolation to microbial diversification. *FEMS Micro Ecol* **48**293 - 303.

**Randa, M. A., Polz, M. F. and Lim, E.** (2004). Effects of Temperature and Salinity on Vibrio vulnificus Population Dynamics as Assessed by Quantitative PCR. *Appl. Environ. Microbiol.* **70**(9), 5469-5476.

**Rasgon, J., Cornel, A. and Scott, T.** (2006). Evolutionary history of a mosquito endosymbiont revealed through mitochondrial hitchhiking. *Proceedings of the Royal Society B: Biological Sciences* **273**(1594), 1603-1611.

**Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. and Whittam, T. S.** (2000). Parallel evolution of virulence in pathogenic *Escherichia coli. Nature* **406**(6791), 64-67.

**Riemann, L. and Azam, F.** (2002). Widespread N-Acetyl-D-Glucosamine Uptake among Pelagic Marine Bacteria and Its Ecological Implications. *Appl. Environ. Microbiol.* **68**(11), 5554-5562.

**Robinson, D. A., Monk, A. B., Cooper, J. E., Feil, E. J. and Enright, M. C.** (2005). Evolutionary Genetics of the Accessory Gene Regulator (*agr*) Locus in *Staphylococcus aureus. J. Bacteriol.* **187**(24), 8312-8321.

**Rogul, M., Brendle, J. J., Haapala, D. K. and Alexander, A. D.** (1970). Nucleic acid similarities among *Pseudomonas pseudomallei*, *Pseudomonas multivorans*, and *Actinobacillus mallei*. *J. Bacteriol.* **101**(3), 827 - 835.

**Roque, A., Molina-Aja, A., Bolan-Mejia, C. and Gomez-Gil, B.** (2001). In vitro susceptibility to 15 antibiotics of vibrios isolated from penaeid shrimps in Northwestern Mexico. *International Journal of Antimicrobial Agents* **17**(5), 383-387.

**Rousset, F., Bouchon, D., Pintureau, B., Juchault, P. and Solignac, M.** (1992a). *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. *Proc. Biol.Sci.* **250**(1328), 91 - 98.

**Rousset, F. and de Stordeur, E.** (1994). Properties of *Drosophila simulans* strains experimentally infected by different clones of the bacterium *Wolbachia*. *Heredity* **72**(4), 325 -331.

**Rousset, F., Vautrin, D. and Solignac, M.** (1992b). Molecular identification of *Wolbachia*, the agent of cytoplasmic incompatibility in *Drosophila simulans*, and variability in relation with host mitochondrial types. *Proc. R. Soc. Lond. B* **247**163-168.

**Rose, T., Schultz, E., Henikoff, J., Pietrokovski, S., McCallum, C. and Henikoff, S.** (1998). Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucl. Acids Res.* **26**(7), 1628-1635.

**Rowley, S. M., Raven, R. J. and McGraw, E. A.** (2004). *Wolbachia pipientis* in Australian Spiders. *Current Microbiology* **49**(3), 208-214.

**Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. and Rozas, R.** (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**(18), 2496-2497.

**Saez, A. G., Tatarenkov, A., Barrio, E., Becerra, N. H. and Ayala, F. J.** (2003). Patterns of DNA sequence polymorphism at *Sod* vicinities in *Drosophila melanogaster*: Unraveling the footprint of a recent selective sweep. *PNAS* **100**(4), 1793-1798.

**Sallstrom, B. and Andersson, S. G.** (2005). Genome reduction in the [alpha]-Proteobacteria. *Current Opinion in Microbiology* **8**(5), 579-585.

**Sammataro, D., Gerson, U. and Needham, G.** (2000). Parasitic Mites of Honey Bees: Life History, Implications, and Impact. *Annual Review of Entomology* **45**(1), 519-548.

**Sanogo, Y. O. and Dobson, S. L.** (2004). Molecular discrimination of Wolbachia in the Culex pipiens complex: Evidence for variable bacteriophage hyperparasitism. *Insect Mol. Biol.* **13**365-369.

**Santos, S. R. and Ochman, H.** (2004). Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environmental Microbiology* **6**(7), 754-759.

**Sawabe, T., Setogushi, N., Inoue, S., Tanaka, R., Ootsubo, M., Yoshimizu, M. and Ezura, Y.** (2003). Acetic acid production of *Vibrio halioticoli* from alginate: a possible role for establishment of abalone-*V. halioticoli* association. *Aquaculture* **219**671-679.

**Sawyer, S.** (1989). Statistical tests for detecting gene conversion. *Mol Biol Evol* **6**(5), 526-538.

**Schell, M. A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M.-C., Desiere, F., Bork, P., Delley, M., Pridmore, R. D and Arigoni, F.** (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *PNAS* **99**(22), 14422-14427.

**Schilthuizen, M. and Stouthamer, R.** (1997). Horizontal transmission of parthenogenesis-inducing microbes in *Trichogramma* wasps. *Proc. R. Soc. Lond. B* **264**(1380), 361-366.

**Schoolnik, G. K. and Yildiz, F. H.** (2000). The complete genome sequence of *Vibrio cholerae*: a tale of two chromosomes and of two lifestyles. *Genome Biology* **1**(3), reviews1016.1 - reviews1016.3.

**Schuch, R. and Fischetti, V. A.** (2006). Detailed Genomic Analysis of the Wβ and γ Phages Infecting *Bacillus anthracis*: Implications for Evolution of Environmental Fitness and Antibiotic Resistance. *J. Bacteriol.* **188**(8), 3037-3051.

**Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. and Whittam, T. S.** (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol.* **51**(5), 873 - 874.

**Selifonova O., Valle F. and Schellenberger V.** (2001) Rapid Evolution of Novel Traits in Microorganisms. *Appl Environ. Microbiol.* **67**(8); 3645-3649.

**Sharp, P. M.** (1997). In search of molecular darwinism. *Nature* **385**(6612), 111-112.

**Sherr, E. B. and Sherr, B. F.** (2002). Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek* **81**(1 - 4), 293-308.

**Sherratt, D.** (2001). Bacteria rule!: The start of a new era in bacterial microbiology? *EMBO reports* **2**(3), 175 - 176.

**Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. and Ishikawa, H.** (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. **407**(6800), 81-86.

**Shrivastava, J., Gower, C. M., Balolong, E., Wang, T. P., Qain, B. Z. and Webster, J. P.** (2005). Population genetics of multi-host parasites – the case for molecular epidemiological studies of *Schistosoma japonicum* using larval stages from naturally infected hosts. *Parasitology* **1311** - 10.

**Sifri C.D., Begun J., Ausubel F.M. and Calderwood S.B.** (2003) *Caenorhabditis elegans* as a Model Host for Staphylococcus aureus Pathogenesis. Infect Immun. **71**(4); 2208-2217.

**Silva, F. J., Latorre, A. and Moya, A.** (2001). Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends in Genetics* **17**(11), 615-618.

**Singleton, F. L., Attwell, R., Jangi, S. and Colwell, R. R.** (1982). Effects of temperature and salinity on *Vibrio cholerae* growth. *Appl. Environ. Microbiol.* **44**(5), 1047-1058.

**Sintupachee, S., Milne, J., Poonchaisri, S., Baimai, V. and Kittayapong, P.** (2006). Closely Related *Wolbachia* Strains within the Pumpkin Arthropod Community and the Potential for Horizontal Transmission via the Plant. *Microbial Ecology*1-8.

**Smith, J., Smith, N., O'Rourke, M. and Spratt, B.** (1993). How Clonal are Bacteria? *PNAS* **90**(10), 4384-4388.

**Smith, J. M., Feil, E. J. and Smith, N. H.** (2000). Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays* **22**(12), 1115-1122.

**Smith, N., Maynard Smith, J. and Spratt, B.** (1995). Sequence evolution of the *porB* gene of *Neisseria gonorrhoeae* and *Neisseria meningitidis*: evidence of positive Darwinian selection. *Mol Biol Evol* **12**(3), 363-370.

**Smyth, D. S., Hartigan, P. J., Meaney, W. J., Fitzgerald, J. R., Deobald, C. F., Bohach, G. A. and Smyth, C. J.** (2005). Superantigen genes encoded by the *egc* cluster and SaPIbov are predominant among *Staphylococcus aureus* isolates from cows, goats, sheep, rabbits and poultry. *J Med Microbiol* **54**(4), 401-411.

**Snel, B., Bork, P. and Huynen, M. A.** (2002). Genomes in Flux: The Evolution of Archaeal and Proteobacterial Gene Content. *Genome Res.* **12**(1), 17-25.

**Sniegowski P.D., Gerrish P.J., Johnson T. and Shaver A.** (2000). The evolution of mutation rates: separating causes from consequences. *Bioessays* **22**; 1057-1066.

**Sorensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. and Wuertz, S.** (2005). Studying plasmid horizontal transfer *in situ*: a critical review. *Nat Rev Micro* **3**(9), 700-710.

**Spratt, B. G., Hanage, W. P. and Feil, E. J.** (2001). The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Current Opinion in Microbiology* **4**(5), 602-606.

**Spratt, B. G., Hanage, W. P., Li, B., Aanensen, D. M. and Feil, E. J.** (2004). Displaying the relatedness among isolates of bacterial species - the eBURST approach. *FEMS Microbiology Letters* **241**(2), 129-134.

**Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P., Kampfer, P., Maiden, M., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H. G., Vauterin, L., Ward, A. C. and Whitman, W. B.** (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**(3), 1043-1047.

**Staley, J. T. and Gosink, J. J.** (1999). POLES APART: Biodiversity and Biogeography of Sea Ice Bacteria. *Annual Review of Microbiology* **53**(1), 189-215.

**Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E.V. and Davis, R.W.** (1998). Genome Sequence of an Obligate Intracellular Pathogen of Humans: *Chlamydia trachomatis*. *Science* **282**(5389), 754-759.

**Stine, O. C., Sozhamannan, S., Gou, Q., Zheng, S., Morris, J. G., Jr. and Johnson, J. A.** (2000). Phylogeny of *Vibrio cholerae* Based on *recA* Sequence. *Infect. Immun.* **68**(12), 7180-7185.

**Storch, D.** (2001). Biodiversity. In *Encyclopedia of Life support Systems*, vol. Chapter 1.1.6.3 (ed.: UNESCO.

**Stouthamer, R., Breeuwer, J. A. J. and Hurst, G. D. D.** (1999). *Wolbachia pipientis*: Microbial Manipulator of Arthropod Reproduction. *Annual Review of Microbiology* **53**(1), 71-102.

**Stouthamer, R., Breeuwer, J. A. J., Luck, R. F. and Werren, J. H.** (1993b). Molecular identification of microorganisms associated with parthenogenesis. *Nature* **361**(6407), 66-68.

**Stouthamer, R. and Werren, J. H.** (1993a). Microbes associated with parthenogenesis in wasps of the genus *Trichogramma*. *J. Invert. Pathol.*(69), 6 - 9.

**Swofford, D. L.** (2000). Phylogenetic analysis using parsimony: Paup* and other methods. Sunderland, Mass.: Sinauer Associates.

**Takeda, K. and Akira, S.** (2003). Toll receptors and pathogen resistance. *Cellular Microbiology* **5**(3), 143-153.

**Tamas, I., Klasson, L., Canback, B., Naslund, A. K., Eriksson, A.-S., Wernegreen, J. J., Sandstrom, J. P., Moran, N. A. and Andersson, S. G. E.** (2002). 50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. *Science* **296**(5577), 2376-2379.

**Tamura, K., Nei, M. and Kumar, S.** (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *PNAS* **101**(30), 11030-11035.

**Taylor, M. J.** (2002). *Wolbachia* Endosymbiotic Bacteria of Filarial Nematodes. A New Insight into Disease Pathogenesis and Control. *Archives of Medical Research* **33**(4), 422-424.

**Taylor, M. J.** (2004). *Wolbachia* host-symbiont interactions. In *Microbe-vector interactions in vector-borne diseases*, vol. 1 (ed. S. H. Gillespie G. L. Smith and A. Osbourn), pp. 287 - 299. Cambridge: Cambridge University Press.

**Terrell, D.** (2003). "*Actinia equina*" (On-line), Animal Diversity Web., vol. 2006 (ed.

**Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G.** (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*(25), 4876 - 4882.

**Thompson, C. C., Thompson, F. L., Vandemeulebroecke, K., Hoste, B., Dawyndt, P. and Swings, J.** (2004a). Use of *recA* as an alternative phylogenetic marker in the family Vibrionaceae. *Int J Syst Evol Microbiol* **54**(3), 919-924.

**Thompson, F. L., Iida, T. and Swings, J.** (2004b). Biodiversity of Vibrios. *Microbiol. Mol. Biol. Rev.* **68**(3), 403-431.

**Thompson, J. R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D. E., Benoit, J., Sarma-Rupavtarm, R., Distel, D. L. and Polz, M. F.** (2005). Genotypic Diversity Within a Natural Coastal Bacterioplankton Population. *Science* **307**(5713), 1311-1313.

**Thompson, J. R., Randa, M. A., Marcelino, L. A., Tomita-Mitchell, A., Lim, E. and Polz, M. F.** (2004c). Diversity and Dynamics of a North Atlantic Coastal *Vibrio* Community. *Appl. Environ. Microbiol.* **70**(7), 4103-4110.

**Tison, D. L., Nishibuchi, M., Greenwood, J. D. and Seidler, R. J.** (1982). *Vibrio vulnificus* biogroup 2: new biogroup pathogenic for eels. *Appl. Environ. Microbiol.* **44**(3), 640-646.

**Tram, U. and Sullivan, W.** (2002). Role of Delayed Nuclear Envelope Breakdown and Mitosis in *Wolbachia*-Induced Cytoplasmic Incompatibility. *Science* **296**(5570), 1124-1126.

**Turelli, M. and Hoffmann, A. A.** (1991). Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature* **353**440-442.

**Turner, P. C., McLennan, A. G., Bates, A. D. and White, M. R. H.** (2001). Instant Notes: Molecular Biology. Oxford, UK: BIOS Scientific Publishers Ltd.

**Vala, F., Egas, M., Breeuwer, J. A. J. and Sabelis, M. W.** (2004). *Wolbachia* affects oviposition and mating behaviour of its spider mite host. *Journal of Evolutionary Biology* **17**(3), 692-700.

**Valvano, M. A., Keith, K. E. and Cardona, S. T.** (2005). Survival and persistence of opportunistic *Burkholderia* species in host cells. *Current Opinion in Microbiology* **8**(1), 99-105.

**Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. and Swings, J.** (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60**(2), 407 - 438.

**van Meer, M. M. M., van Kan, F. J. P. M. and Stouthamer, R.** (1999a). Spacer 2 region and 5S rDNA variation of *Wolbachia* strains involved in cytoplasmic incompatibility or sex-ratio distortion in arthropods. *Letters in Applied Microbiology* **28**(1), 17-22.

**van Meer, M. M. M., Witteveldt, J. and Stouthamer, R.** (1999b). Phylogeny of the arthropod endosymbiont *Wolbachia* based on the *wsp* gene. *Insect Molecular Biology* **8**(3), 399-408.

**Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D., Levy, S., Knap, A., Lomas, M., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. and Smith, H.** (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304**(5667), 66-74.

**Verdonck, L., Grisez, L., Sweetman, E., Minkoff, G., Sorgeloos, P., Ollevier, F. and Swings, J.** (1997). Vibrios associatedwith routine productions of *Brachionus plicatilis*. *Aquaculture* **149**203-214.

**Verschuere, L., Rombaut, G., Sorgeloos, P. and Verstraete, W.** (2000). Probiotic Bacteria as Biological Control Agents in Aquaculture. *Microbiol. Mol. Biol. Rev.* **64**(4), 655-671.

**Vugia, D., Shefer, A., Douglas, J., Greene, K., Bryant, R. and Werner, S.** (1997). Cholera from raw seaweed transported from the Philippines to California. *J. Clin. Microbiol.* **35**(1), 284-285.

**Wagner, A.** (2001). The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes. *Mol Biol Evol* **18**(7), 1283-1292.

**Walter, D. E., Krantz, G. and Lindquist, E.** (1996). Acari - The Mites, vol. 2006 (ed.: The Tree of Life Web Project).

**Waterfield, N. R., Wren, B. W. and ffrench-Constant, R. H.** (2004). Invertebrates as a source of emerging human pathogens. *Nat Rev Micro* **2**(10), 833-841.

**Watterson, G. A.** (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**(2), 256-276.

**Weeks, A. R. and Breeuwer, J. A. J.** (2001). *Wolbachia* induced parthenogenesis in a genus of phytophagous mites. *Proceedings of the Royal Society B: Biological Sciences* **268**(1482), 2245-2251.

**Weller, T. M. A.** (2000). Methicillin-resistant *Staphylococcus aureus* typing methods: which should be the international standard? *Journal of Hospital Infection* **44**(3), 160-172.

**Wernegreen, J. and Moran, N.** (1999). Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. *Mol Biol Evol* **16**(1), 83-97.

**Werren, J. H.** (1997a). Biology of *Wolbachia. Annu. Rev. Entomol.* **42**587-609.

**Werren, J. H.** (1997b). *Wolbachia* run amok. *PNAS* **94**(21), 11154-11155.

**Werren, J. H. and Bartos, J. D.** (2001). Recombination in *Wolbachia. Current Biology* **11**(6), 431-435.

**Werren, J. H., Windsor, D. M. and Guo, L. R.** (1995a). Distribution of *Wolbachia* among neotropical arthropods. *Proc. R. Soc. Lond. B*(262), 197 - 204.

**Werren, J. H., Zhang, W. and Guo, L. R.** (1995b). Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc. Biol.Sci.* **261**(1360), 53 - 63.

**White, J. C. and Liburd, O. E.** (2005). Effects of soil moisture and temperature on reproduction and development of twospotted spider mite (Acari: Tetranychidae) in strawberries. *J Econ Entemol.* **98**(1), 154 - 158.

**Whitaker, R. J., Grogan, D. W. and Taylor, J. W.** (2003). Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Science* **301**(5635), 976-978.

**Whitman, W. B., Coleman, D. C. and Wiebe, W. J.** (1998). Prokaryotes: The unseen majority. *PNAS* **95**(12), 6578-6583.

**Whittam, T., Ochman, H. and Selander, R.** (1983). Geographic components of linkage disequilibrium in natural populations of *Escherichia coli. Mol Biol Evol* **1**(1), 67-83.

**Winzer, K. and Williams, P.** (2001). Quorum sensing and the regulation of virulence gene expression in pathogenic bacteria. *Int. J. Med. Microbiol.* **291**131 - 143.

**Woese, C. R.** (1991). The Use of Ribosomal RNA in Reconstructing Evolutionary Relationships Among Bacteria. In *Evolution at the Molecular Level*, (ed. R. Selander A. G. Clark and T. Whittam), pp. 1 - 24. Sunderland: Sinauer Associates Inc.

**Wolfe, K. H. and Shields, D. C.** (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**(6634), 708-713.

**Woolhouse, M. E. J., Taylor, L. H. and Haydon, D. T.** (2001). Population Biology of Multihost Pathogens. *Science* **292**(5519), 1109-1112.

**Woo Oh, H., Gwang Kim, M., Woon Shin, S., Sook Bae, K., Joon Ahn, Y. and Park, H.-Y.** (2000). Ultrastructural and molecular identification of a *Wolbachia* endosymbiont in a spider, *Nephila clavata. Insect Molecular Biology* **9**(5), 539-543.

**Wren, B. W.** (2000). Microbial genome analysis: insights into virulence, host adaptation and evolution. *Nat Rev Genet* **1**(1), 30-39.

**Wright, A., Hill, R., Johnson, J., Roghman, M., Colwell, R. and Morris, J., Jr.** (1996). Distribution of *Vibrio vulnificus* in the Chesapeake Bay. *Appl. Environ. Microbiol.* **62**(2), 717-724.

**Wright, S.** (1951). The genetical structure of populations. *Annals of Eugenics*(15), 323 - 354.

**Wu, M., Sun, L. V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J. C., McGraw, E. A., Martin, W., Esser, C., Ahmadinejad, N., Wiegand, C., Madupu, R., Beanan, M. J., Brinkac, L. M., Daugherty, S. C., Durkin, A. S., Kolonay, J. F., Nelson, W. C., Mohamoud, Y., Lee, P., Berry, K., Young, M. B., Utterback, T., Weidman, J., Nierman, W. C., Paulsen, I. T., Nelson, K. E., Tettelin, H., O'Neill, S. L. and Eisen, J A.** (2004). Phylogenomics of the reproductive parasite *Wolbachia pipientis* *w*Mel: a streamlined genome overrun by mobile genetic elements. *PLOS Biol.* **2**E69-E83.

**Yang, Z., Goldman, N. and Friday, A.** (1994). Comparison of models for nucleotide substitution used in maximum- likelihood phylogenetic estimation. *Mol Biol Evol* **11**(2), 316-324.

**Yen, J. H. and Barr, A. R.** (1971). New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature*(232), 657 - 658.

**Young, J. P. W., Downer, H. L. and Eardly, B. D.** (1991). Phylogeny of the Phototrophic Rhizobium Strain BTAi1 by Polymerase Chain Reaction-Based Sequencing of a 16S rRNA Gene Segment. *Journal of Bacteriology* **173**(7), 2271-2277.

**Yu X.-Q., Zhu Y.-F., Ma C., Fabrick J.A. and Kanost M.R.** (2002) Pattern recognition proteins in *Manduca sexta* plasma. *Insect Biochem. Mol Biol.* **32**; 1287-1293.

**Zchori-Fein, E. and Perlman, S. J.** (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular Ecology* **13**(7), 2009-2016.

**Zhang, Z. and Kishino, H.** (2004). Genomic Background Predicts the Fate of Duplicated Genes: Evidence From the Yeast Genome. *Genetics* **166**(4), 1995-1999.

**Zhou, W., Rousset, F. and O'Neill, S. L.** (1998). Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proc. R. Soc. Lond. B* **265** 509-515.

# APPENDIX A

## SUPPLEMENTARY INFORMATION FOR CHAPTER THREE

Details of arthropod hosts used in this study (n=44)

| ID | Order | Family | Species name | Location & collection date |
|---|---|---|---|---|
| **A01** | Hymenoptera | Formicidae | Unknown | Unknown |
| **B03** | Coleoptera | Tenebrionidae | *Tribolium confusum* | Silwood, UK, 2002 |
| **B04** | Coleoptera | Tenebrionidae | *Tribolium confusum* | Unknown |
| **B05** | Coleoptera | Tenebrionidae | *Tribolium castaneum* | Unknown |
| **BEE01** | Hymenoptera | Apidae | Unknown | Silwood, UK |
| **BEE02** | Hymenoptera | Apidae | Unknown | Silwood, UK |
| **BF01** | Lepidoptera | Hesperiidae | Unknown | Silwood, UK, 2000 |
| **F01** | Diptera | Drosophilidae | *Drosophila simulans/melanogaster hybrid* | Unknown |
| **F02** | Diptera | Drosophilidae | *Drosophila simulans* | Dundee lab culture |
| **F03** | Diptera | Drosophilidae | *Drosophila subobscura* | Silwood, UK, 2002 |
| **F04** | Diptera | Drosophilidae | *Drosophila simulans* | Silwood lab culture |
| **F05** | Diptera | Drosophilidae | *Drosophila simulans* | Cyprus lab culture |
| **HF01** | Diptera | Syrphidae | Unknown | Unknown |
| **M02** | Lepidoptera | Plutellidae | *Plutella xylostella* | Cornwall, UK, 2002 |
| **BF02** | Lepidoptera | Pieridae | *Pieris rapae* | Cornwall, UK, 2002 |
| **M04** | Lepidoptera | Choreutidae | *Anthophila fabriciana* | Silwood, UK, 2003 |
| **M05** | Lepidoptera | Noctuidae | *Helicoverpa armigera* | Unknown |
| **M06** | Lepidoptera | Noctuidae | *Agrotis ipsilon* | Switzerland, 2000 |
| **M07** | Lepidoptera | Choreutidae | *Anthophila fabriciana* | Loch Tay, Scotland, 2002 |
| **M08** | Lepidoptera | Choreutidae | *Anthophila fabriciana* | Bath, UK, 2002 |
| **M09** | Lepidoptera | Choreutidae | *Anthophila fabriciana* | Leeds, UK, 2002 |
| **M10** | Lepidoptera | Plutellidae | *Plutella xylostella* | Unknown |
| **M13** | Lepidoptera | Plutellidae | *Plutella xylostella* | Unknown |
| **M14** | Lepidoptera | Plutellidae | *Plutella xylostella* | Unknown |
| **M17** | Lepidoptera | Plutellidae | *Plutella xylostella* | Pahang, Malaysia, 2001 |
| **M18** | Lepidoptera | Plutellidae | *Plutella xylostella* | Pahang, Malaysia, 2001 |
| **M22** | Lepidoptera | Plutellidae | *Plutella xylostella* | Sungei Mensum, Malaysia, 2001 |

| | | | | |
|---|---|---|---|---|
| **M25** | Lepidoptera | Choreutidae | *Anthophila fabriciana* | Unknown |
| **PW01** | Hymenoptera | Braconidae | *Macrocentrus grandii* | Newark, USA, 1996 |
| **PW03** | Hymenoptera | Ichneumonidae | *Bracon hebetor* | lab culture, Dundee, Scotland |
| **PW04** | Hymenoptera | Ichneumonidae | *Phytodietus niger* | Costa Rica |
| **PW07** | Hymenoptera | Braconidae | *Asobara tabida* | Silwood lab culture |
| **PW08** | Hymenoptera | Ichneumonidae | *Campoletis sonorensis* | Leeds, UK, 2002 |
| **PW09** | Hymenoptera | Unknown | Unknown | Unknown |
| **PW14** | Hymenoptera | Braconidae | *Asobara tabida* | Unknown |
| **PW15** | Hymenoptera | Pteromalidae | *Roptrocerus xylophagorum* | Germany |
| **SEF01** | Diptera | Diopsidae | *Olescampe mascellator* | Unknown |
| **SEF02** | Diptera | Diopsidae | *Cyrtodiopsis dalmanni* | Unknown |
| **SEF03** | Diptera | Diopsidae | Unknown | Unknown |
| **SEF04** | Diptera | Diopsidae | *Cyrtodiopsis whitei* | Thailand, 2004 |
| **SF01** | Hymenoptera | Diprionidae | *Diprion similis* | UK |
| **SM26** | Acari | Tetranychidae | *Bryobia sarothamni* | Amsterdam, Netherlands, 2004 |
| **T01** | Isoptera | Unknown | Unknown | Thailand, 2004 |
| **W02** | Hymenoptera | Vespidae | *Polistes fuscatus* | Switzerland, 2002 |

*Wolbachia* spp. primers

***wsp***
For 5' TGG TCC AAT ARR TGA TGA AGA AAC TAG CTA 3'
Rev 5' AAA AAT TAA ACG YTA CTC CAG CTT CYG CAC3'

***ftsZ***
For 5' GTT GTC GCA AAT ACY GAT GC 3'
Rev 5' CTT AAR TAA GCT GGT ATA TC 3'

***atpA***
For 5' GGT GCC TGT AGG GCA TGA ATT 3'
Rev 5' GCT AAA AAT TGC ATA GGC GCG CAG 3'

***gltA***
For 5' GAY CAT GAR CAR AAY GCT TCT AC 3'
Rev 5' CCH GAR TAA AAA TCA ACR TTD GG 3'

***groEL***
For 5' CAA CRG TRG SRR YAA CTG CDG G 3'
Rev 5' GAT ATC CRC GRT CAA AYT GC 3'

***trmD***
For 5' GAA CTA TTC TCT TTG CCG GAA AAG 3'
Rev 5' CAC TGC TCA GGT CTA GTA TAT TGA GG 3'

***tpiA***
For 5' GGA ATG CGT TCT TCA TTT GTT GAC 3'
Rev 5' CCT GTG CCT ATT GCC C 3'

nt-ST & aa-ST cross reference table

| Strain | nt-ST | aa-ST |
|---|---|---|
| **A01** | 1 | 1 |
| **B03** | 2 | 2 |
| **B04** | 23 | 3 |
| **B05** | 23 | 3 |
| **BEE01** | 3 | 4 |
| **BEE02** | 4 | 30 |
| **BF01** | 24 | 21 |
| **BF02** | 5 | 23 |
| **F01** | 28 | 5 |
| **F02** | 30 | 4 |
| **F03** | 29 | 31 |
| **F04** | 28 | 4 |
| **F05** | 6 | 4 |
| **HF01** | 31 | 6 |
| **M02** | 25 | 22 |
| **M04** | 7 | 7 |
| **M05** | 8 | 8 |
| **M06** | 24 | 21 |
| **M07** | 26 | 25 |
| **M08** | 9 | 1 |
| **M09** | 42 | 9 |
| **M10** | 31 | 26 |
| **M13** | 36 | 10 |
| **M14** | 37 | 32 |
| **M17** | 33 | 27 |
| **M18** | 34 | 28 |
| **M22** | 35 | 29 |
| **M25** | 11 | 9 |
| **PW01** | 12 | 11 |
| **PW03** | 13 | 12 |
| **PW04** | 38 | 33 |
| **PW07** | 14 | 4 |
| **PW08** | 15 | 13 |
| **PW09** | 16 | 14 |
| **PW14** | 17 | 15 |
| **PW15** | 18 | 35 |
| **SEF01** | 19 | 16 |
| **SEF02** | 20 | 17 |
| **SEF03** | 39 | 18 |
| **SEF04** | 21 | 34 |
| **SF01** | 27 | 19 |
| **SM26** | 22 | 24 |
| **T01** | 40 | 20 |
| **W02** | 41 | 36 |

# APPENDIX B

## SUPPLEMENTARY INFORMATION FOR CHAPTER FOUR

Details of arthropod hosts used in this study (n=20)

| ID | Order | Family | Species | Host plant | Date of isolation | Location |
|---|---|---|---|---|---|---|
| **SM01** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 19/04/2004 | Alcossebres, Spain |
| **SM04** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 03/04/2004 | Amsterdam, NL |
| **SM06** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 06/04/2004 | Hardegarijp, NL |
| **SM07** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 10/05/2004 | Angers, France |
| **SM09** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 04/05/2004 | Valkenburg, NL |
| **SM10** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 19/04/2004 | Gatova, Spain |
| **SM11** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 26/05/2004 | Lompret, Belgium |
| **SM12** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 20/04/2004 | Alcossebres, Spain |
| **SM13** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 15/04/2004 | Varsseveld, NL |
| **SM14** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 14/04/2004 | Begues, Spain |
| **SM15** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 19/04/2004 | Alcossebres, Spain |
| **SM16** | Acari | Tetranychidae | *Bryobia praetiosa* | *Vicia* sp. | 26/05/2004 | Vierves s Viroin, Belgium |
| **SM17** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 03/04/2004 | Amsterdam, NL |
| **SM19** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 06/04/2004 | Hardegarijp, NL |
| **SM20** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 10/05/2004 | Angers, France |
| **SM21** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 04/05/2004 | Varsseveld, NL |
| **SM22** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 19/04/2004 | Gatova, Spain |
| **SM23** | Acari | Tetranychidae | *Bryobia kissophila* | *Hedera helix* | 26/05/2004 | Lompret, Belgium |
| **SM26** | Acari | Tetranychidae | *Bryobia sarothamni* | *Cytisus scoparius* | 26/05/2004 | Vierves s Viroin, Belgium |
| **SM28** | Acari | Tetranychidae | *Tetranychus urticae* | *Rosmarinus ? Lavandula ?* | ??/05/2004 | Valencia, Spain |

nt-ST & aa-ST cross reference table

| Strain | nt-ST | aa-ST |
|---|---|---|
| SM01 | 1 | 1 |
| SM04 | 12 | 7 |
| SM06 | 7 | 8 |
| SM07 | 7 | 8 |
| SM09 | 8 | 9 |
| SM10 | 2 | 2 |
| SM11 | 12 | 7 |
| SM12 | 3 | 12 |
| SM13 | 9 | 10 |
| SM14 | 4 | 4 |
| SM15 | 2 | 2 |
| SM16 | 10 | 11 |
| SM17 | 7 | 8 |
| SM19 | 7 | 8 |
| SM20 | 11 | 5 |
| SM21 | 7 | 8 |
| SM22 | 2 | 2 |
| SM23 | 7 | 8 |
| SM26 | 5 | 6 |
| SM28 | 6 | 6 |

# APPENDIX C

## SUPPLEMENTARY INFORMATION FOR CHAPTER FIVE

*Vibrio* spp. primers

**16S rRNA**
For 5' TGG CTC AGA ACG AAC GCT GGC GGC 3'
Rev 5' CCC ACT GCT GCC TCC CGT AGG AGT 3'

***recA***
For 5' TGG ACG AGA ATA AAC AGA AGG C 3'
Rev 5' CCG TTA TAG CTG TAC CAA GCG CCC 3'

***recG***
For 5' GGG CGA CGT GGG CDS NGG NAA RAC 3'
Rev 5' GGG TCC GGG GGA TNG GNG TNG C 3'

***mdh***
For 5' GAT CTG AGY CAT ATC CCW AC 3'
Rev 5' GCT TCW ACM ACY TCR GTA CCC G 3'

***ompK***
For 5' AAY GAT TAC AAS TGG ATG CAA TT 3'
Rev 5' GAR TGC CAG TAR AKA CCG TT 3'

Details of *Vibrio* spp. used in this study (n=164)

| Strain | Source | Date | ST | 16SrRNA | recG | ompK | mdh | recA | Genbank ID |
|---|---|---|---|---|---|---|---|---|---|
| S1A65iii | *A. equina* | 06/04 | 2 | 9 | 2 | 2 | 2 | 2 | *V. cyclitrophicus* |
| S1A65ix | *A. equina* | 06/04 | 6 | 9 | 4 | 4 | 2 | 2 | *V. cyclitrophicus* |
| S1A65v | *A. equina* | 06/04 | 18 | 9 | 5 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1A65vi | *A. equina* | 06/04 | 24 | 9 | 6 | 5 | 3 | 2 | *V. cyclitrophicus* |
| S1A65vii | *A. equina* | 06/04 | 6 | 9 | 4 | 4 | 2 | 2 | *V. cyclitrophicus* |
| S1A65x | *A. equina* | 06/04 | 6 | 9 | 4 | 4 | 2 | 2 | *V. cyclitrophicus* |
| S1A66i | *A. equina* | 06/04 | 26 | 9 | 7 | 6 | 2 | 3 | *V. cyclitrophicus* |
| S1A66ii | *A. equina* | 06/04 | 28 | 9 | 8 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1A66iii | *A. equina* | 06/04 | 3 | 9 | 2 | 2 | 2 | 4 | *V. cyclitrophicus* |
| S1A66iv | *A. equina* | 06/04 | 31 | 9 | 9 | 5 | 2 | 4 | *V. cyclitrophicus* |
| S1C01i | *C. maenas* | 06/04 | 12 | 9 | 4 | 5 | 2 | 5 | *V. cyclitrophicus* |
| S1C01ii | *C. maenas* | 06/04 | 8 | 9 | 4 | 5 | 2 | 4 | *V. cyclitrophicus* |
| S1C01iii | *C. maenas* | 06/04 | 22 | - | 5 | 7 | 3 | 2 | *V. cyclitrophicus* |
| S1C01iv | *C. maenas* | 06/04 | 12 | 9 | 4 | 5 | 2 | 5 | *V. cyclitrophicus* |
| S1L01ii | *P. vulgata* | 06/04 | 33 | 9 | 10 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1L01iii | *P. vulgata* | 06/04 | 35 | 9 | 10 | 5 | 2 | 6 | *V. cyclitrophicus* |
| S1L01iv | *P. vulgata* | 06/04 | 19 | 9 | 5 | 5 | 2 | 7 | *V. cyclitrophicus* |
| S1L01vi | *P. vulgata* | 06/04 | 8 | 9 | 4 | 5 | 2 | 4 | *V. cyclitrophicus* |
| S1L02i | *P. vulgata* | 06/04 | 17 | 9 | 5 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S1L02iii | *P. vulgata* | 06/04 | 126 | 9 | 12 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1L02v | *P. vulgata* | 06/04 | 43 | - | 14 | 10 | 5 | 4 | *V. cyclitrophicus* |
| S1M01i | *M. edulis* | 06/04 | 11 | 9 | 4 | 5 | 2 | 8 | *V. cyclitrophicus* |
| S1M01ii | *M. edulis* | 06/04 | 46 | 9 | 15 | 5 | 6 | 3 | *V. cyclitrophicus* |
| S1M01iii | *M. edulis* | 06/04 | 18 | 9 | 5 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1M01iv | *M. edulis* | 06/04 | 125 | 9 | 10 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S1M01v | *M. edulis* | 06/04 | 44 | 9 | 15 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S1M02i | *M. edulis* | 06/04 | 48 | 9 | 16 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1M02iii | *M. edulis* | 06/04 | 18 | 9 | 5 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1M02iv | *M. edulis* | 06/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1M02v | *M. edulis* | 06/04 | 20 | 9 | 5 | 5 | 7 | 3 | *V. cyclitrophicus* |
| S1S18i | Sea water | 06/04 | 50 | - | 18 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S1S18ii | Sea water | 06/04 | 50 | 9 | 18 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S1S18iii | Sea water | 06/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S1S18iv | Sea water | 06/04 | 37 | 9 | 10 | 6 | 2 | 4 | *V. cyclitrophicus* |
| S1S18v | Sea water | 06/04 | 48 | 9 | 16 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA02vi | *A. equina* | 07/04 | 51 | 10 | 20 | 13 | 2 | 11 | *V. kanaloae* |
| S2aA02vii | *A. equina* | 07/04 | 8 | 9 | 4 | 5 | 2 | 4 | *V. cyclitrophicus* |
| S2aA02viii | *A. equina* | 07/04 | 18 | 9 | 5 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA03i | *A. equina* | 07/04 | 52 | - | 21 | 14 | 11 | 13 | *V. kanaloae* |
| S2aA03ii | *A. equina* | 07/04 | 53 | 10 | 21 | 14 | 12 | 13 | *V. kanaloae* |
| S2aA03vi | *A. equina* | 07/04 | 54 | 10 | 22 | 16 | 13 | 14 | *V. kanaloae* |
| S2aA03vii | *A. equina* | 07/04 | 54 | 10 | 22 | 16 | 13 | 14 | *V. kanaloae* |
| S2aA04i | *A. equina* | 07/04 | 48 | 9 | 16 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA04ii | *A. equina* | 07/04 | 33 | 9 | 10 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA04iii | *A. equina* | 07/04 | 55 | 16 | 23 | 19 | 16 | 17 | *V. kanaloae* |
| S2aA04iv | *A. equina* | 07/04 | 56 | 18 | 24 | 20 | 17 | 13 | *V. kanaloae* |
| S2aA04vi | *A. equina* | 07/04 | 57 | 22 | 25 | 21 | 18 | 18 | *V. tasmaniensis* |
| S2aA04vii | *A. equina* | 07/04 | 57 | 22 | 25 | 21 | 18 | 18 | *V. tasmaniensis* |
| S2aA04x | *A. equina* | 07/04 | 62 | 9 | 26 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA05ix | *A. equina* | 07/04 | 64 | 22 | 27 | 22 | 19 | 19 | *V. tasmaniensis* |
| S2aA05vi | *A. equina* | 07/04 | 63 | 22 | 27 | 22 | 19 | 3 | *V. cyclitrophicus* |
| S2aA05x | *A. equina* | 07/04 | 59 | 11 | 25 | 23 | 20 | 3 | *V. cyclitrophicus* |

| Strain | Source | Date | ST | 16SrRNA | recG | ompK | mdh | recA | Genbank ID |
|---|---|---|---|---|---|---|---|---|---|
| S2aA06ix | *A. equina* | 07/04 | 21 | 9 | 5 | 5 | 21 | 2 | *V. cyclitrophicus* |
| S2aA06vi | *A. equina* | 07/04 | 21 | 9 | 5 | 5 | 21 | 2 | *V. cyclitrophicus* |
| S2aA06vii | *A. equina* | 07/04 | 21 | 9 | 5 | 5 | 21 | 2 | *V. cyclitrophicus* |
| S2aA06viii | *A. equina* | 07/04 | 65 | 31 | 28 | 5 | 22 | 20 | *V. splendidus* |
| S2aA06x | *A. equina* | 07/04 | 21 | 9 | 5 | 5 | 21 | 2 | *V. cyclitrophicus* |
| S2aA09vi | *A. equina* | 07/04 | 38 | 10 | 10 | 25 | 2 | 21 | *V. splendidus* |
| S2aA09vii | *A. equina* | 07/04 | 33 | 9 | 10 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA09viii | *A. equina* | 07/04 | 17 | 9 | 5 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2aA09x | *A. equina* | 07/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA10iii | *A. equina* | 07/04 | 58 | 10 | 25 | 23 | 20 | 23 | *V. tasmaniensis* |
| S2aA11ix | *A. equina* | 07/04 | 62 | 9 | 26 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA13ii | *A. equina* | 07/04 | 68 | 9 | 32 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2aA13iv | *A. equina* | 07/04 | 70 | 9 | 33 | 28 | 25 | 26 | *V. kanaloae* |
| S2aA13v | *A. equina* | 07/04 | 71 | 9 | 34 | 29 | 2 | 3 | *V. cyclitrophicus* |
| S2aA14i | *A. equina* | 07/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA14iv | *A. equina* | 07/04 | 72 | 9 | 35 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA28i | *A. equina* | 07/04 | 73 | 14 | 36 | 21 | 26 | 27 | *V. tasmaniensis* |
| S2aA28v | *A. equina* | 07/04 | 75 | 9 | 37 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA29ix | *A. equina* | 07/04 | 76 | 10 | 38 | 23 | 26 | 28 | *V. kanaloae* |
| S2aA29vi | *A. equina* | 07/04 | 77 | 27 | 39 | 30 | 27 | 29 | *V. pomeroyi* |
| S2aA29vii | *A. equina* | 07/04 | 78 | 16 | 41 | 18 | 28 | 30 | *V. splendidus* |
| S2aA29x | *A. equina* | 07/04 | 78 | 16 | 41 | 18 | 28 | 30 | *V. splendidus* |
| S2aA30ix | *A. equina* | 07/04 | 79 | 9 | 42 | 12 | 29 | 30 | *V. splendidus* |
| S2aA30vi | *A. equina* | 07/04 | 80 | 16 | 43 | 32 | 10 | 30 | *V. splendidus* |
| S2aA30vii | *A. equina* | 07/04 | 17 | 9 | 5 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2aA30viii | *A. equina* | 07/04 | 81 | 26 | 44 | 23 | 30 | 19 | *V. tasmaniensis* |
| S2aA30x | *A. equina* | 07/04 | 82 | 22 | 45 | 23 | 26 | 31 | *V. kanaloae* |
| S2aA31i | *A. equina* | 07/04 | 84 | 16 | 46 | 23 | 22 | 32 | *V. splendidus* |
| S2aA31ii | *A. equina* | 07/04 | 85 | 10 | 46 | 23 | 31 | 27 | *V. tasmaniensis* |
| S2aA31iii | *A. equina* | 07/04 | 58 | 13 | 25 | 23 | 20 | 23 | *V. tasmaniensis* |
| S2aA32ix | *A. equina* | 07/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA32vi | *A. equina* | 07/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA32vii | *A. equina* | 07/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA32viii | *A. equina* | 07/04 | 7 | 9 | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2aA32x | *A. equina* | 07/04 | 65 | 32 | 28 | 5 | 22 | 20 | *V. splendidus* |
| S2aA34viii | *A. equina* | 07/04 | 90 | 16 | 48 | 20 | 22 | 32 | *V. splendidus* |
| S2aA34x | *A. equina* | 07/04 | 91 | 9 | 49 | 35 | 25 | 33 | *V. kanaloae* |
| S2aA67ap | *A. equina* | 07/04 | 93 | 3 | 50 | 36 | 24 | 1 | *V. harveyi* |
| S2aS01iii | Sea water | 07/04 | 94 | 10 | 51 | 37 | 32 | 34 | *V. kanaloae* |
| S2bA41viii | *A. equina* | 07/04 | 58 | 18 | 25 | 23 | 20 | 23 | *V. tasmaniensis* |
| S2bA41x | *A. equina* | 07/04 | 61 | 10 | 25 | 31 | 34 | 31 | *V. kanaloae* |
| S2bA42i | *A. equina* | 07/04 | 34 | 9 | 4 | 6 | 2 | 36 | *V. cyclitrophicus* |
| S2bA42ii | *A. equina* | 07/04 | 69 | 9 | 32 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA42v | *A. equina* | 07/04 | 13 | 9 | 4 | 7 | 2 | 3 | *V. cyclitrophicus* |
| S2bA44vii | *A. equina* | 07/04 | 89 | 16 | 47 | 38 | 10 | 20 | *V. tasmaniensis* |
| S2bA44viii | *A. equina* | 07/04 | 25 | 9 | 6 | 42 | 2 | 20 | *V. splendidus* |
| S2bA44x | *A. equina* | 07/04 | 68 | 9 | 32 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2bA46x | *A. equina* | 07/04 | 95 | 10 | 54 | 43 | 36 | 39 | *V. kanaloae* |
| S2bA48iii | *A. equina* | 07/04 | 74 | 21 | 36 | 44 | 26 | 40 | *V. kanaloae* |
| S2bA48iv | *A. equina* | 07/04 | 97 | 1 | 55 | 45 | 37 | 30 | *V. splendidus* |
| S2bA48v | *A. equina* | 07/04 | 47 | 9 | 15 | 6 | 2 | 41 | *V. pomeroyi* |
| S2bA48vi | *A. equina* | 07/04 | 60 | 10 | 25 | 23 | 26 | 42 | *V. tasmaniensis* |
| S2bA48vii | *A. equina* | 07/04 | 32 | 9 | 10 | 4 | 2 | 2 | *V. cyclitrophicus* |
| S2bA48viii | *A. equina* | 07/04 | 98 | 10 | 56 | 46 | 39 | 43 | *V. splendidus* |
| S2bA48x | *A. equina* | 07/04 | 92 | 9 | 49 | 47 | 40 | 26 | *V. kanaloae* |

| Strain | Source | Date | ST | 16SrRNA | recG | ompK | mdh | recA | Genbank ID |
|---|---|---|---|---|---|---|---|---|---|
| S2bA50vii | *A. equina* | 07/04 | 9 | - | 4 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2bA50viii | *A. equina* | 07/04 | 7 | - | 4 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA51iii | *A. equina* | 07/04 | 99 | 10 | 57 | 48 | 2 | 44 | *V. splendidus* |
| S2bA51iv | *A. equina* | 07/04 | 4 | 9 | 2 | 4 | 2 | 3 | *V. cyclitrophicus* |
| S2bA52vi | *A. equina* | 07/04 | 33 | 9 | 10 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA52viii | *A. equina* | 07/04 | 14 | 9 | 4 | 42 | 2 | 47 | *V. cyclitrophicus* |
| S2bA53vi | *A. equina* | 07/04 | 60 | 10 | 25 | 23 | 26 | 42 | *V. tasmaniensis* |
| S2bA53vii | *A. equina* | 07/04 | 33 | 9 | 10 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA53viii | *A. equina* | 07/04 | 83 | 10 | 45 | 51 | 42 | 19 | *V. tasmaniensis* |
| S2bA53x | *A. equina* | 07/04 | 60 | 10 | 25 | 23 | 26 | 42 | *V. tasmaniensis* |
| S2bA57i | *A. equina* | 07/04 | 29 | 10 | 8 | 52 | 2 | 24 | *V. pomeroyi* |
| S2bA57iii | *A. equina* | 07/04 | 36 | 10 | 10 | 5 | 25 | 48 | *V. kanaloae* |
| S2bA57vi | *A. equina* | 07/04 | 5 | 9 | 2 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2bA57vii | *A. equina* | 07/04 | 28 | 9 | 8 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA58ii | *A. equina* | 07/04 | 10 | 9 | 4 | 5 | 2 | 6 | *V. cyclitrophicus* |
| S2bA58iii | *A. equina* | 07/04 | 17 | 9 | 5 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2bA59ii | *A. equina* | 07/04 | 100 | 37 | 61 | 23 | 20 | 20 | *V. splendidus* |
| S2bA61v | *A. equina* | 07/04 | 48 | 9 | 16 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA62i | *A. equina* | 07/04 | 39 | 9 | 10 | 56 | 47 | 52 | *V. kanaloae* |
| S2bA62ii | *A. equina* | 07/04 | 60 | 10 | 25 | 23 | 26 | 42 | *V. tasmaniensis* |
| S2bA62iii | *A. equina* | 07/04 | 40 | 9 | 10 | 57 | 2 | 26 | *V. kanaloae* |
| S2bA62iv | *A. equina* | 07/04 | 60 | 10 | 25 | 23 | 26 | 42 | *V. tasmaniensis* |
| S2bA62vi | *A. equina* | 07/04 | 101 | 10 | 63 | 23 | 30 | 18 | *V. tasmaniensis* |
| S2bA62viii | *A. equina* | 07/04 | 9 | 9 | 4 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2bA63ix | *A. equina* | 07/04 | 102 | 10 | 64 | 58 | 48 | 53 | *V. kanaloae* |
| S2bA63vi | *A. equina* | 07/04 | 96 | 10 | 54 | 59 | 48 | 54 | *V. kanaloae* |
| S2bA64vi | *A. equina* | 07/04 | 125 | 9 | 10 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S2bA64vii | *A. equina* | 07/04 | 28 | 9 | 8 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA64x | *A. equina* | 07/04 | 28 | 9 | 8 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S2bA70ap | *A. equina* | 07/04 | 104 | 4 | 68 | 1 | 1 | 1 | *V. harveyi* |
| S3A72ii | *A. equina* | 05/05 | 105 | 10 | 69 | 61 | 52 | 60 | *V. splendidus* |
| S3A75ii | *A. equina* | 05/05 | 86 | 29 | 46 | 52 | 63 | 67 | *V. splendidus* |
| S3A75v | *A. equina* | 05/05 | 108 | 29 | 71 | 65 | 65 | 61 | *V. splendidus* |
| S3A75vi | *A. equina* | 05/05 | 107 | 30 | 71 | 31 | 36 | 61 | *V. splendidus* |
| S3A77v | *A. equina* | 05/05 | 87 | 16 | 47 | 15 | 72 | 32 | *V. splendidus* |
| S3A77vii | *A. equina* | 05/05 | 88 | - | 47 | 27 | 60 | 20 | *V. splendidus* |
| S3A79vi | *A. equina* | 05/05 | 109 | 16 | 72 | 69 | 28 | 20 | *V. splendidus* |
| S3L05i | *P. vulgata* | 05/05 | 110 | 42 | 74 | 31 | 76 | 30 | *V. splendidus* |
| S4aA80ii | *A. equina* | 08/05 | 111 | 10 | 76 | 23 | 30 | 77 | *V. tasmaniensis* |
| S4aA80iii | *A. equina* | 08/05 | 112 | 10 | 76 | 72 | 30 | 77 | *V. tasmaniensis* |
| S4aA80iv | *A. equina* | 08/05 | 41 | 9 | 10 | 73 | 2 | 3 | *V. cyclitrophicus* |
| S4aA81i | *A. equina* | 08/05 | 30 | 10 | 8 | 74 | 77 | 78 | *V. tasmaniensis* |
| S4aA81iii | *A. equina* | 08/05 | 33 | 9 | 10 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S4aA81v | *A. equina* | 08/05 | 111 | 10 | 76 | 23 | 30 | 77 | *V. tasmaniensis* |
| S4aP10iv | *Littorina* spp. | 08/05 | 23 | 9 | 5 | 80 | 2 | 3 | *V. cyclitrophicus* |
| S4aP10v | *Littorina* spp. | 08/05 | 9 | 9 | 4 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S4aS09i | Sea water | 08/05 | 27 | 9 | 8 | 4 | 2 | 3 | *V. cyclitrophicus* |
| S4aS09ii | Sea water | 08/05 | 16 | 9 | 4 | 82 | 2 | 3 | *V. cyclitrophicus* |
| S4bA84ii | *A. equina* | 08/05 | 42 | 9 | 10 | 86 | 2 | 3 | *V. cyclitrophicus* |
| S4bA84iii | *A. equina* | 08/05 | 122 | 22 | 89 | 87 | 89 | 91 | *V. lentus* |
| S4bA84v | *A. equina* | 08/05 | 122 | 22 | 89 | 87 | 89 | 91 | *V. lentus* |
| S4bP14iv | *Littorina* spp. | 08/05 | 58 | 10 | 25 | 23 | 20 | 23 | *V. tasmaniensis* |
| S4bP14v | *Littorina* spp. | 08/05 | 123 | 10 | 92 | 88 | 93 | 34 | *V. kanaloae* |
| S4bP16ii | *Littorina* spp. | 08/05 | 45 | 9 | 15 | 5 | 2 | 3 | *V. cyclitrophicus* |
| S4bP18i | *Littorina* spp. | 08/05 | 15 | - | 4 | 80 | 2 | 3 | *V. cyclitrophicus* |

| Strain | Source | Date | ST | 16SrRNA | recG | ompK | mdh | recA | Genbank ID |
|---|---|---|---|---|---|---|---|---|---|
| S4bP18iii | *Littorina* spp. | 08/05 | 125 | 9 | 10 | 5 | 2 | 2 | *V. cyclitrophicus* |
| S4bS14i | Sea water | 08/05 | 124 | 9 | 96 | 91 | 40 | 52 | *V. kanaloae* |

30 randomly selected *Vibrio* isolates

| Strain | *recA* species assignment |
|---|---|
| S1S18iii | *V. cyclitrophicus* |
| S2bA51iv | *V. cyclitrophicus* |
| S1A65ix | *V. cyclitrophicus* |
| S2b52vi | *V. cyclitrophicus* |
| S1A66iv | *V. cyclitrophicus* |
| S4bP16ii | *V. cyclitrophicus* |
| S3A72ii | *V. splendidus* |
| S2aA67ap | *V. harveyi* |
| S4bS14i | *V. kanaloae* |
| S1L02v | *V. cyclitrophicus* |
| S4bA84v | *V. lentus* |
| S2aA32x | *V. splendidus* |
| S2aA29x | *V. splendidus* |
| S3A77v | *V. splendidus* |
| S2bA48iii | *V. kanaloae* |
| S2aA05ix | *V. tasmaniensis* |
| S2aS01iii | *V. kanaloae* |
| S2bA48x | *V. kanaloae* |
| S2bA63ix | *V. kanaloae* |
| S2aA03ii | *V. kanaloae* |
| S2A75v | *V. splendidus* |
| S3L05i | *V. splendidus* |
| S2bA59ii | *V. splendidus* |
| S4bP14iv | *V. tasmaniensis* |
| S2aA30viii | *V. tasmaniensis* |
| S2bA53vi | *V. tasmaniensis* |
| S2aA04iii | *V. kanaloae* |
| S2bA48v | *V. pomeroyi* |
| S2aA81i | *V. tasmaniensis* |
| S2aA29vi | *V. pomeroyi* |

nt-ST & aa-ST cross reference table

| Strain | nt-ST | aa-ST |
|---|---|---|
| S1A65iii | 2 | 15 |
| S1A65ix | 6 | 23 |
| S1A65v | 18 | 15 |
| S1A65vi | 24 | 16 |
| S1A65vii | 6 | 23 |
| S1A65x | 6 | 23 |
| S1A66i | 26 | 22 |
| S1A66ii | 28 | 15 |
| S1A66iii | 3 | 15 |
| S1A66iv | 31 | 17 |
| S1C01i | 12 | 15 |
| S1C01ii | 8 | 15 |
| S1C01iii | 22 | 23 |
| S1C01iv | 12 | 15 |
| S1L01ii | 33 | 15 |
| S1L01iii | 35 | 61 |
| S1L01iv | 19 | 62 |
| S1L01vi | 8 | 15 |
| S1L02i | 17 | 15 |
| S1L02iii | 126 | 15 |
| S1L02v | 43 | 56 |
| S1M01i | 11 | 15 |
| S1M01ii | 46 | 15 |
| S1M01iii | 18 | 15 |
| S1M01iv | 125 | 15 |
| S1M01v | 44 | 15 |
| S1M02i | 48 | 15 |
| S1M02iii | 18 | 15 |
| S1M02iv | 7 | 15 |
| S1M02v | 20 | 35 |
| S1S18i | 50 | 15 |
| S1S18ii | 50 | 15 |
| S1S18iii | 7 | 15 |
| S1S18iv | 37 | 22 |
| S1S18v | 48 | 15 |
| S2aA02vi | 51 | 19 |
| S2aA02vii | 8 | 15 |
| S2aA02viii | 18 | 15 |
| S2aA03i | 52 | 8 |
| S2aA03ii | 53 | 8 |
| S2aA03vi | 54 | 52 |
| S2aA03vii | 54 | 5 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S2aA04i | 48 | 15 |
| S2aA04ii | 33 | 15 |
| S2aA04iii | 55 | 7 |
| S2aA04iv | 56 | 33 |
| S2aA04vi | 57 | 1 |
| S2aA04vii | 57 | 1 |
| S2aA04x | 62 | 15 |
| S2aA05ix | 64 | 11 |
| S2aA05vi | 63 | 11 |
| S2aA05x | 59 | 2 |
| S2aA06ix | 21 | 27 |
| S2aA06vi | 21 | 27 |
| S2aA06vii | 21 | 27 |
| S2aA06viii | 65 | 40 |
| S2aA06x | 21 | 27 |
| S2aA09vi | 38 | 14 |
| S2aA09vii | 33 | 15 |
| S2aA09viii | 17 | 15 |
| S2aA09x | 7 | 15 |
| S2aA10iii | 58 | 2 |
| S2aA11ix | 62 | 15 |
| S2aA13ii | 68 | 15 |
| S2aA13iv | 70 | 30 |
| S2aA13v | 71 | 15 |
| S2aA14i | 7 | 15 |
| S2aA14iv | 72 | 15 |
| S2aA28i | 73 | 1 |
| S2aA28v | 75 | 18 |
| S2aA29ix | 76 | 3 |
| S2aA29vi | 77 | 20 |
| S2aA29viii | 78 | 43 |
| S2aA29x | 78 | 43 |
| S2aA30ix | 79 | 58 |
| S2aA30vi | 80 | 44 |
| S2aA30vii | 17 | 15 |
| S2aA30viii | 81 | 2 |
| S2aA30x | 82 | 3 |
| S2aA31i | 84 | 42 |
| S2aA31ii | 85 | 37 |
| S2aA31iii | 58 | 63 |
| S2aA32ix | 7 | 15 |
| S2aA32vi | 7 | 15 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S2aA32vii | 7 | 15 |
| S2aA32viii | 7 | 15 |
| S2aA32x | 65 | 40 |
| S2aA34viii | 90 | 41 |
| S2aA34x | 91 | 29 |
| S2aA67ap | 93 | 53 |
| S2aAS01iii | 94 | 6 |
| S2bA41vii | 89 | 46 |
| S2bA41viii | 58 | 2 |
| S2bA41x | 61 | 26 |
| S2bA42i | 34 | 22 |
| S2bA42ii | 69 | 15 |
| S2bA42v | 13 | 23 |
| S2bA44viii | 25 | 39 |
| S2bA44x | 68 | 15 |
| S2bA46x | 95 | 10 |
| S2bA48iii | 74 | 13 |
| S2bA48iv | 97 | 49 |
| S2bA48v | 47 | 22 |
| S2bA48vi | 60 | 2 |
| S2bA48vii | 32 | 23 |
| S2bA48viii | 98 | 34 |
| S2bA48x | 92 | 31 |
| S2bA50ii | 9 | 15 |
| S2bA50iii | 7 | 15 |
| S2bA51iii | 99 | 57 |
| S2bA51iv | 4 | 23 |
| S2bA52vi | 33 | 15 |
| S2bA52viii | 14 | 59 |
| S2bA53vi | 60 | 2 |
| S2bA53vii | 33 | 15 |
| S2bA53viii | 83 | 12 |
| S2bA53x | 60 | 64 |
| S2bA57i | 29 | 21 |
| S2bA57iii | 36 | 28 |
| S2bA57vi | 5 | 15 |
| S2bA57vii | 28 | 15 |
| S2bA58ii | 10 | 60 |
| S2bA58iii | 17 | 15 |
| S2bA59ii | 100 | 38 |
| S2bA61v | 48 | 15 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S2bA62i | 39 | 28 |
| S2bA62ii | 60 | 2 |
| S2bA62iii | 40 | 15 |
| S2bA62iv | 60 | 2 |
| S2bA62vi | 101 | 3 |
| S2bA62viii | 9 | 15 |
| S2bA63ix | 102 | 4 |
| S2bA63vi | 96 | 9 |
| S2bA64vi | 125 | 15 |
| S2bA64vii | 28 | 15 |
| S2bA64x | 28 | 15 |
| S2bA70ap | 104 | 54 |
| S3A72ii | 105 | 55 |
| S3A75ii | 86 | 50 |
| S3A75v | 108 | 41 |
| S3A75vi | 107 | 26 |
| S3A77v | 87 | 45 |
| S3A77vii | 88 | 48 |
| S3A79i | 109 | 47 |
| S3L05i | 110 | 51 |
| S4aA80ii | 111 | 2 |
| S4aA80iii | 112 | 2 |
| S4aA80iv | 41 | 23 |
| S4aA81i | 30 | 36 |
| S4aA81iii | 33 | 15 |
| S4aA81v | 111 | 2 |
| S4aP10iv | 23 | 15 |
| S4aP10v | 9 | 15 |
| S4aS09i | 27 | 23 |
| S4aS09ii | 16 | 15 |
| S4bA84ii | 42 | 24 |
| S4bA84iii | 122 | 25 |
| S4bA84v | 122 | 25 |
| S4bP14iv | 58 | 2 |
| S4bP14v | 123 | 6 |
| S4bP16ii | 45 | 15 |
| S4bP18i | 15 | 15 |
| S4bP18iii | 125 | 15 |
| S4bS14i | 124 | 32 |

# APPENDIX D

## SUPPLEMENTARY INFORMATION FOR CHAPTER SIX

Details of New Zealand *Vibrio* spp. used in this study (n=105)

| Strain | Date | Location | Source | Genbank ID |
|---|---|---|---|---|
| 1 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 2 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 4 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 5 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 7 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 12 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. harveyi* |
| 13 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. chagasii* |
| 14 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. fortis* |
| 15 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. chagasii* |
| 16 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 17 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 18 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 19 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 20 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. campbellii* |
| 21 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 23 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 25 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 28 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. chagasii* |
| 30 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 32 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. cincinnatiensis* |
| 33 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. harveyi* |
| 34 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. harveyi* |
| 35 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 36 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 38 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 39 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. kanaloae* |
| 41 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 42 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 43 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. harveyi* |
| 44 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 45 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 46 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. harveyi* |
| 47 | 22/02/2005 | Bethells Beach | *A. tenebrosa* | *V. pomeroyi* |
| 51 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 52 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 53 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 54 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. tasmaniensis* |
| 56 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 57 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. kanaloae* |
| 58 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. kanaloae* |
| 59 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. kanaloae* |

| Strain | Date | Location | Source | Genbank ID |
|---|---|---|---|---|
| 60 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. kanaloae* |
| 61 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 62 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 63 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. kanaloae* |
| 64 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. harveyi* |
| 66 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 67 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. diabolicus* |
| 68 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. diabolicus* |
| 72 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. diabolicus* |
| 76 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 77 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. harveyi* |
| 78 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 80 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. fortis* |
| 82 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 83 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 84 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. pomeroyi* |
| 91 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. kanaloae* |
| 93 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. fortis* |
| 95 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. kanaloae* |
| 96 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. campbellii* |
| 98 | 22/02/2005 | O'Neills Beach | *A. tenebrosa* | *V. harveyi* |
| 100 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 101 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 103 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 105 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. harveyi* |
| 106 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *P. damselae* |
| 108 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. campbellii* |
| 109 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. chagasii* |
| 112 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. fortis/chagasii* |
| 113 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *P. damselae* |
| 130 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 133 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. harveyi* |
| 134 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. harveyi* |
| 137 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 141 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 142 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 143 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 145 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 146 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. pomeroyi* |
| 149 | 24/02/2005 | Narrow Neck Beach | *A. tenebrosa* | *V. harveyi* |
| 154 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. fortis* |
| 158 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. alginolyticus* |
| 161 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. campbellii* |
| 162 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. harveyi* |

| Strain | Date | Location | Source | Genbank ID |
|---|---|---|---|---|
| 163 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. pomeroyi* |
| 164 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. pomeroyi* |
| 166 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. pomeroyi* |
| 172 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. harveyi* |
| 176 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. fortis* |
| 178 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. harveyi* |
| 180 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. harveyi* |
| 186 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. pomeroyi* |
| 187 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. campbellii* |
| 188 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. harveyi* |
| 189 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. pomeroyi* |
| 190 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. campbellii* |
| 191 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. pomeroyi* |
| 192 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. pomeroyi* |
| 194 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. harveyi* |
| 195 | 24/02/2005 | Westwell Beach | *A. tenebrosa* | *V. fortis/chagasii* |
| 199 | 22/02/2005 | O'Neills Beach | Sea water | *V. harveyi* |
| 200 | 22/02/2005 | O'Neills Beach | Sea water | *V. harveyi* |
| 202 | 22/02/2005 | O'Neills Beach | Sea water | *V. campbellii* |
| 206 | 22/02/2005 | O'Neills Beach | Sea water | *V. harveyi* |

Details of UK *Vibrio* spp. used in this study (n=180).

| Strain | Source | Date | Genbank ID |
|---|---|---|---|
| S1A65i | *A. equina* | 06/04 | *V. harveyi* |
| S1A65iii | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A65ix | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A65v | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A65vi | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A65vii | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A65x | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A66i | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A66ii | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A66iii | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1A66iv | *A. equina* | 06/04 | *V. cyclitrophicus* |
| S1C01i | *C. maenas* | 06/04 | *V. cyclitrophicus* |
| S1C01ii | *C. maenas* | 06/04 | *V. cyclitrophicus* |
| S1C01iii | *C. maenas* | 06/04 | *V. cyclitrophicus* |
| S1C01iv | *C. maenas* | 06/04 | *V. cyclitrophicus* |
| S1L01ii | *P. vulgata* | 06/04 | *V. cyclitrophicus* |
| S1L01iii | *P. vulgata* | 06/04 | *V. cyclitrophicus* |
| S1L01iv | *P. vulgata* | 06/04 | *V. cyclitrophicus* |
| S1L01vi | *P. vulgata* | 06/04 | *V. cyclitrophicus* |
| S1L02i | *P. vulgata* | 06/04 | *V. cyclitrophicus* |
| S1L02iii | *P. vulgata* | 06/04 | *V. cyclitrophicus* |
| S1L02v | *P. vulgata* | 06/04 | *V. cyclitrophicus* |
| S1M01i | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M01ii | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M01iii | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M01iv | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M01v | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M02i | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M02iii | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M02iv | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1M02v | *M. edulis* | 06/04 | *V. cyclitrophicus* |
| S1P01i | *Littorina* spp. | 06/04 | *V. diabolicus* |
| S1P01ii | *Littorina* spp. | 06/04 | *V. diabolicus* |
| S1S18i | Sea water | 06/04 | *V. cyclitrophicus* |
| S1S18ii | Sea water | 06/04 | *V. cyclitrophicus* |
| S1S18iii | Sea water | 06/04 | *V. cyclitrophicus* |
| S1S18iv | Sea water | 06/04 | *V. cyclitrophicus* |
| S1S18v | Sea water | 06/04 | *V. cyclitrophicus* |
| S2aA02vi | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA02vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA02viii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA03i | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA03ii | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA03vi | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA03vii | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA04i | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA04ii | *A. equina* | 07/04 | *V. cyclitrophicus* |

| Strain | Source | Date | Genbank ID |
|---|---|---|---|
| S2aA04iii | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA04iv | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA04vi | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA04vii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA04x | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA05ix | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA05vi | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA05x | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA06ix | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA06vi | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA06vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA06viii | *A. equina* | 07/04 | *V. splendidus* |
| S2aA06x | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA09ix | *A. equina* | 07/04 | *V. splendidus* |
| S2aA09vi | *A. equina* | 07/04 | *V. splendidus* |
| S2aA09vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA09viii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA09x | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA10iii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA11ix | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA13i | *A. equina* | 07/04 | *V. harveyi* |
| S2aA13ii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA13iv | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA13v | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA14i | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA14iv | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA28i | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA28v | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA29ix | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA29vi | *A. equina* | 07/04 | *V. pomeroyi* |
| S2aA29vii | *A. equina* | 07/04 | *V. splendidus* |
| S2aA29x | *A. equina* | 07/04 | *V. splendidus* |
| S2aA30ix | *A. equina* | 07/04 | *V. splendidus* |
| S2aA30vi | *A. equina* | 07/04 | *V. splendidus* |
| S2aA30vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA30viii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA30x | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA31i | *A. equina* | 07/04 | *V. splendidus* |
| S2aA31ii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA31iii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2aA32ix | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA32vi | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA32vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA32viii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2aA32x | *A. equina* | 07/04 | *V. splendidus* |
| S2aA34viii | *A. equina* | 07/04 | *V. splendidus* |
| S2aA34x | *A. equina* | 07/04 | *V. kanaloae* |
| S2aA67ap | *A. equina* | 07/04 | *V. harveyi* |
| S2aS01iii | Sea water | 07/04 | *V. kanaloae* |

| Strain | Source | Date | Genbank ID |
|---|---|---|---|
| S2bA41viii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA41x | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA42i | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA42ii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA42v | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA44vii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA44viii | *A. equina* | 07/04 | *V. splendidus* |
| S2bA44x | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA46x | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA48iii | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA48iv | *A. equina* | 07/04 | *V. splendidus* |
| S2bA48v | *A. equina* | 07/04 | *V. pomeroyi* |
| S2bA48vi | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA48vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA48viii | *A. equina* | 07/04 | *V. splendidus* |
| S2bA48x | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA50vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA50viii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA51iii | *A. equina* | 07/04 | *V. splendidus* |
| S2bA51iv | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA52vi | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA52viii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA53vi | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA53vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA53viii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA53x | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA57i | *A. equina* | 07/04 | *V. pomeroyi* |
| S2bA57iii | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA57vi | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA57vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA58ii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA58iii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA59ii | *A. equina* | 07/04 | *V. splendidus* |
| S2bA61v | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA62i | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA62ii | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA62iii | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA62iv | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA62vi | *A. equina* | 07/04 | *V. tasmaniensis* |
| S2bA62viii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA63ix | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA63vi | *A. equina* | 07/04 | *V. kanaloae* |
| S2bA64vi | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA64vii | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA64x | *A. equina* | 07/04 | *V. cyclitrophicus* |
| S2bA69ap | *A. equina* | 07/04 | *V. diabolicus* |
| S2bA70ap | *A. equina* | 07/04 | *V. harveyi* |
| S3A72ii | *A. equina* | 05/05 | *V. splendidus* |
| S3A73i | *A. equina* | 05/05 | *V. splendidus* |

| Strain | Source | Date | Genbank ID |
|---|---|---|---|
| S3A75ii | *A. equina* | 05/05 | *V. splendidus* |
| S3A75v | *A. equina* | 05/05 | *V. splendidus* |
| S3A75vi | *A. equina* | 05/05 | *V. splendidus* |
| S3A77v | *A. equina* | 05/05 | *V. splendidus* |
| S3A77vii | *A. equina* | 05/05 | *V. splendidus* |
| S3A79vi | *A. equina* | 05/05 | *V. splendidus* |
| S3L05i | *P. vulgata* | 05/05 | *V. splendidus* |
| S4aA80ii | *A. equina* | 08/05 | *V. tasmaniensis* |
| S4aA80iii | *A. equina* | 08/05 | *V. tasmaniensis* |
| S4aA80iv | *A. equina* | 08/05 | *V. cyclitrophicus* |
| S4aA81i | *A. equina* | 08/05 | *V. tasmaniensis* |
| S4aA81iii | *A. equina* | 08/05 | *V. cyclitrophicus* |
| S4aA81v | *A. equina* | 08/05 | *V. tasmaniensis* |
| S4aP04iii | *Littorina* spp. | 08/05 | *V. diabolicus* |
| S4aP08i | *Littorina* spp. | 08/05 | *V. diabolicus* |
| S4aP10ii | *Littorina* spp. | 08/05 | *V. diabolicus* |
| S4aP10iii | *Littorina* spp. | 08/05 | *V. diabolicus* |
| S4aP10iv | *Littorina* spp. | 08/05 | *V. cyclitrophicus* |
| S4aP10v | *Littorina* spp. | 08/05 | *V. cyclitrophicus* |
| S4aP16v | *Littorina* spp. | 08/05 | *V. diabolicus* |
| S4aS05v | Sea water | 08/05 | *V. diabolicus* |
| S4aS09i | Sea water | 08/05 | *V. cyclitrophicus* |
| S4aS09ii | Sea water | 08/05 | *V. cyclitrophicus* |
| S4aS09iv | Sea water | 08/05 | *V. diabolicus* |
| S4aS09v | Sea water | 08/05 | *V. diabolicus* |
| S4bA84ii | *A. equina* | 08/05 | *V. cyclitrophicus* |
| S4bA84iii | *A. equina* | 08/05 | *V. lentus* |
| S4bA84v | *A. equina* | 08/05 | *V. lentus* |
| S4bP14iv | *Littorina* spp. | 08/05 | *V. tasmaniensis* |
| S4bP14v | *Littorina* spp. | 08/05 | *V. kanaloae* |
| S4bP16ii | *Littorina* spp. | 08/05 | *V. cyclitrophicus* |
| S4bP18i | *Littorina* spp. | 08/05 | *V. cyclitrophicus* |
| S4bP18iii | *Littorina* spp. | 08/05 | *V. cyclitrophicus* |
| S4bS11i | Sea water | 08/05 | *V. diabolicus* |
| S4bS14i | Sea water | 08/05 | *V. kanaloae* |

Nucleotide sequence types and allelic profiles (n=285)

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| S2aA09ix | 1 | 1 | 1 | 54 |
| S2aA09vi | 2 | 1 | 2 | 54 |
| S1A65v | 3 | 2 | 2 | 55 |
| S1A66ii | 3 | 2 | 2 | 55 |
| S1L01ii | 3 | 2 | 2 | 55 |
| S1L02iii | 3 | 2 | 2 | 55 |
| S1M01ii | 3 | 2 | 2 | 55 |
| S1M01iii | 3 | 2 | 2 | 55 |
| S1M02i | 3 | 2 | 2 | 55 |
| S1M02iii | 3 | 2 | 2 | 55 |
| S1M02iv | 3 | 2 | 2 | 55 |
| S1S18iii | 3 | 2 | 2 | 55 |
| S1S18v | 3 | 2 | 2 | 55 |
| S2aA02viii | 3 | 2 | 2 | 55 |
| S2aA04i | 3 | 2 | 2 | 55 |
| S2aA04ii | 3 | 2 | 2 | 55 |
| S2aA04x | 3 | 2 | 2 | 55 |
| S2aA09vii | 3 | 2 | 2 | 55 |
| S2aA09x | 3 | 2 | 2 | 55 |
| S2aA11ix | 3 | 2 | 2 | 55 |
| S2aA13v | 3 | 2 | 2 | 55 |
| S2aA14i | 3 | 2 | 2 | 55 |
| S2aA14iv | 3 | 2 | 2 | 55 |
| S2aA28v | 3 | 2 | 2 | 55 |
| S2aA32ix | 3 | 2 | 2 | 55 |
| S2aA32vi | 3 | 2 | 2 | 55 |
| S2aA32vii | 3 | 2 | 2 | 55 |
| S2aA32viii | 3 | 2 | 2 | 55 |
| S2bA42ii | 3 | 2 | 2 | 55 |
| S2bA52vi | 3 | 2 | 2 | 55 |
| S2bA53vii | 3 | 2 | 2 | 55 |
| S2bA57vii | 3 | 2 | 2 | 55 |
| S2bA61v | 3 | 2 | 2 | 55 |
| S2bA64vii | 3 | 2 | 2 | 55 |
| S2bA64x | 3 | 2 | 2 | 55 |
| S4aA81iii | 3 | 2 | 2 | 55 |
| S4aP10iv | 3 | 2 | 2 | 55 |
| S4aS09ii | 3 | 2 | 2 | 55 |
| S4bP16ii | 3 | 2 | 2 | 55 |
| S4bP18i | 3 | 2 | 2 | 55 |
| S1S18ii | 4 | 2 | 2 | 69 |
| S1A66i | 5 | 2 | 2 | 77 |

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| S2bA42i | 5 | 2 | 2 | 77 |
| S2bA42v | 6 | 2 | 2 | 78 |
| S2bA51iv | 6 | 2 | 2 | 78 |
| S4aS09i | 6 | 2 | 2 | 78 |
| S4aA80iv | 7 | 2 | 2 | 104 |
| S4bA84ii | 8 | 2 | 2 | 106 |
| S2aA05x | 9 | 2 | 16 | 59 |
| S2aA05vi | 10 | 2 | 19 | 71 |
| S1M02v | 11 | 2 | 31 | 55 |
| S1L02i | 13 | 3 | 2 | 55 |
| S1M01iv | 13 | 3 | 2 | 55 |
| S1M01v | 13 | 3 | 2 | 55 |
| S1S18i | 13 | 3 | 2 | 55 |
| S2aA09viii | 13 | 3 | 2 | 55 |
| S2aA13ii | 13 | 3 | 2 | 55 |
| S2aA30vii | 13 | 3 | 2 | 55 |
| S2bA44x | 13 | 3 | 2 | 55 |
| S2bA50vii | 13 | 3 | 2 | 55 |
| S2bA57vi | 13 | 3 | 2 | 55 |
| S2bA58iii | 13 | 3 | 2 | 55 |
| S2bA62viii | 13 | 3 | 2 | 55 |
| S2bA64vi | 13 | 3 | 2 | 55 |
| S4aP10v | 13 | 3 | 2 | 55 |
| S4bP18iii | 13 | 3 | 2 | 55 |
| S1A65ix | 14 | 3 | 2 | 78 |
| S1A65vii | 14 | 3 | 2 | 78 |
| S1A65x | 14 | 3 | 2 | 78 |
| S2bA48vii | 14 | 3 | 2 | 78 |
| S1A65iii | 15 | 3 | 2 | 91 |
| S2aA06ix | 16 | 3 | 4 | 55 |
| S2aA06vi | 16 | 3 | 4 | 55 |
| S2aA06vii | 16 | 3 | 4 | 55 |
| S2aA06x | 16 | 3 | 4 | 55 |
| S1A65vi | 17 | 3 | 22 | 55 |
| S2bA50viii | 17 | 3 | 22 | 55 |
| S1C01iii | 18 | 3 | 22 | 78 |
| S1A66iv | 19 | 4 | 2 | 55 |
| S1C01ii | 19 | 4 | 2 | 55 |
| S1L01vi | 19 | 4 | 2 | 55 |
| S2aA02vii | 19 | 4 | 2 | 55 |
| S1S18iv | 12 | 4 | 2 | 77 |
| S1A66iii | 21 | 4 | 2 | 91 |

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| S2bA44viii | 22 | 5 | 2 | 75 |
| S2aA06viii | 23 | 5 | 3 | 55 |
| S2aA32x | 23 | 5 | 3 | 55 |
| S3A79i | 24 | 5 | 9 | 95 |
| S2bA59ii | 25 | 5 | 16 | 59 |
| S2bA41vii | 26 | 5 | 17 | 88 |
| S3A77vii | 27 | 5 | 17 | 109 |
| S2aA04vi | 28 | 6 | 5 | 56 |
| S2aA04vii | 28 | 6 | 5 | 56 |
| S2bA62vi | 29 | 6 | 18 | 59 |
| S2aA34viii | 30 | 7 | 3 | 57 |
| S2aA31i | 31 | 7 | 3 | 59 |
| S3A77v | 32 | 7 | 35 | 94 |
| S2aA34x | 33 | 8 | 6 | 58 |
| S2aA29ix | 34 | 9 | 7 | 59 |
| S2aA29vi | 35 | 10 | 8 | 60 |
| S2aA30ix | 36 | 11 | 2 | 69 |
| S2aA29viii | 37 | 11 | 9 | 61 |
| S2aA29x | 37 | 11 | 9 | 61 |
| S2aA30vi | 38 | 11 | 17 | 70 |
| S2bA48iv | 39 | 11 | 24 | 83 |
| S3A73i | 40 | 11 | 33 | 93 |
| S3L05i | 41 | 11 | 46 | 89 |
| S2aA03vi | 42 | 12 | 10 | 62 |
| S2aA03vii | 43 | 13 | 10 | 62 |
| S2aS01iii | 44 | 14 | 11 | 63 |
| S4bP14v | 45 | 14 | 41 | 102 |
| S2aA28i | 46 | 15 | 7 | 56 |
| S2aA31ii | 47 | 15 | 48 | 59 |
| S2aA04iii | 48 | 16 | 12 | 64 |
| S2aA04iv | 49 | 17 | 13 | 57 |
| S2aA03ii | 50 | 17 | 14 | 65 |
| S2aA03i | 51 | 17 | 44 | 65 |
| S2aA13i | 52 | 18 | 15 | 66 |
| S2bA62iii | 53 | 19 | 2 | 81 |
| S2aA13iv | 54 | 19 | 6 | 67 |
| S2bA48x | 55 | 19 | 25 | 86 |
| S2aA10iii | 56 | 20 | 16 | 59 |
| S2bA41viii | 56 | 20 | 16 | 59 |
| S4bP14iv | 56 | 20 | 16 | 59 |
| S2aA31iii | 57 | 21 | 16 | 59 |
| 105 | 58 | 22 | 15 | 3 |
| 43 | 172 | 22 | 15 | 11 |
| S2aA67ap | 59 | 22 | 15 | 68 |

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| S1A65i | 60 | 22 | 28 | 66 |
| S2bA70ap | 60 | 22 | 28 | 66 |
| 46 | 61 | 22 | 56 | 11 |
| 134 | 61 | 22 | 56 | 11 |
| 206 | 61 | 22 | 56 | 11 |
| 149 | 62 | 22 | 60 | 11 |
| 77 | 63 | 22 | 60 | 51 |
| 200 | 64 | 22 | 78 | 33 |
| S2bA53viii | 65 | 23 | 7 | 72 |
| S2aA30viii | 66 | 23 | 18 | 59 |
| S2aA05ix | 67 | 23 | 19 | 71 |
| S2aA30x | 68 | 24 | 7 | 59 |
| S2bA41x | 69 | 24 | 29 | 89 |
| S2bA48vi | 70 | 25 | 7 | 59 |
| S2bA53vi | 70 | 25 | 7 | 59 |
| S2bA62ii | 70 | 25 | 7 | 59 |
| S2bA62iv | 70 | 25 | 7 | 59 |
| S2bA53x | 173 | 26 | 7 | 59 |
| S2bA63ix | 71 | 27 | 20 | 73 |
| S2bA63vi | 72 | 28 | 20 | 74 |
| S2bA52viii | 73 | 29 | 2 | 75 |
| S2bA46x | 74 | 30 | 21 | 76 |
| S2bA51iii | 75 | 31 | 2 | 79 |
| S2bA62i | 76 | 32 | 23 | 80 |
| S4bS14i | 77 | 32 | 25 | 98 |
| S2bA48iii | 78 | 33 | 7 | 82 |
| S2bA48v | 79 | 34 | 2 | 77 |
| S2bA57i | 80 | 35 | 2 | 84 |
| 142 | 81 | 35 | 59 | 14 |
| 143 | 81 | 35 | 59 | 14 |
| 66 | 82 | 35 | 93 | 1 |
| S2bA57iii | 83 | 36 | 6 | 55 |
| S2bA58ii | 84 | 37 | 2 | 55 |
| S2bA48viii | 85 | 38 | 26 | 85 |
| S2bA69ap | 86 | 39 | 27 | 87 |
| S1P01i | 87 | 40 | 30 | 90 |
| S1P01ii | 87 | 40 | 30 | 90 |
| S1M01i | 88 | 41 | 2 | 55 |
| S1L01iii | 89 | 42 | 2 | 55 |
| S1L01iv | 90 | 43 | 2 | 55 |
| S1C01i | 91 | 44 | 2 | 55 |
| S1C01iv | 91 | 44 | 2 | 55 |
| S3A72ii | 92 | 45 | 32 | 92 |
| S3A75ii | 93 | 46 | 34 | 84 |

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| S3A75v | 94 | 47 | 9 | 89 |
| S3A75vi | 94 | 47 | 9 | 89 |
| S4aS05v | 95 | 48 | 22 | 96 |
| S4bP16v | 96 | 48 | 47 | 96 |
| 72 | 97 | 48 | 97 | 49 |
| S4aS09iv | 98 | 49 | 36 | 87 |
| S4aS09v | 99 | 49 | 37 | 87 |
| S4bS11i | 100 | 50 | 38 | 97 |
| S4aP04iii | 101 | 51 | 38 | 99 |
| S4aP08i | 102 | 52 | 39 | 100 |
| S4aP10ii | 103 | 53 | 40 | 101 |
| S4aA80ii | 104 | 54 | 18 | 59 |
| S4aA81v | 104 | 54 | 18 | 59 |
| S4aA80iii | 105 | 54 | 18 | 103 |
| S4aA81i | 106 | 55 | 42 | 105 |
| S4bA84iii | 107 | 56 | 43 | 107 |
| S4bA84v | 107 | 56 | 43 | 107 |
| S2aA02vi | 108 | 57 | 2 | 89 |
| S1L02v | 109 | 58 | 45 | 108 |
| S4aP10iii | 110 | 59 | 49 | 90 |
| 1 | 111 | 60 | 50 | 1 |
| 2 | 111 | 60 | 50 | 1 |
| 5 | 111 | 60 | 50 | 1 |
| 61 | 112 | 60 | 50 | 36 |
| 100 | 113 | 61 | 51 | 2 |
| 101 | 113 | 61 | 51 | 2 |
| 103 | 113 | 61 | 51 | 2 |
| 130 | 113 | 61 | 51 | 2 |
| 141 | 113 | 61 | 51 | 2 |
| 145 | 113 | 61 | 51 | 2 |
| 146 | 113 | 61 | 51 | 2 |
| 106 | 114 | 62 | 52 | 4 |
| 113 | 114 | 62 | 52 | 4 |
| 108 | 115 | 63 | 53 | 5 |
| 109 | 116 | 64 | 54 | 6 |
| 13 | 117 | 64 | 57 | 9 |
| 15 | 117 | 64 | 57 | 9 |
| 112 | 118 | 65 | 55 | 7 |
| 195 | 118 | 65 | 55 | 7 |
| 12 | 119 | 66 | 56 | 8 |
| 64 | 120 | 66 | 71 | 27 |
| 133 | 121 | 67 | 15 | 10 |
| 33 | 122 | 67 | 60 | 11 |
| 34 | 122 | 67 | 60 | 11 |

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| 98 | 122 | 67 | 60 | 11 |
| 137 | 123 | 68 | 50 | 12 |
| 14 | 124 | 69 | 58 | 13 |
| 154 | 125 | 70 | 61 | 15 |
| 158 | 126 | 71 | 62 | 16 |
| 16 | 127 | 72 | 63 | 17 |
| 17 | 127 | 72 | 63 | 17 |
| 19 | 127 | 72 | 63 | 17 |
| 161 | 128 | 73 | 64 | 18 |
| 162 | 129 | 74 | 65 | 19 |
| 163 | 130 | 75 | 66 | 20 |
| 164 | 130 | 75 | 66 | 20 |
| 166 | 130 | 75 | 66 | 20 |
| 172 | 131 | 76 | 67 | 21 |
| 176 | 132 | 77 | 68 | 22 |
| 178 | 133 | 78 | 67 | 23 |
| 180 | 133 | 78 | 67 | 23 |
| 18 | 134 | 79 | 50 | 24 |
| 47 | 135 | 79 | 50 | 36 |
| 186 | 136 | 80 | 69 | 25 |
| 187 | 137 | 81 | 70 | 26 |
| 188 | 138 | 82 | 71 | 27 |
| 194 | 139 | 82 | 75 | 11 |
| 189 | 140 | 83 | 72 | 28 |
| 192 | 141 | 83 | 72 | 29 |
| 190 | 142 | 84 | 73 | 29 |
| 191 | 143 | 85 | 74 | 30 |
| 199 | 144 | 86 | 76 | 31 |
| 20 | 145 | 87 | 77 | 32 |
| 96 | 146 | 88 | 79 | 34 |
| 202 | 146 | 88 | 79 | 34 |
| 21 | 147 | 89 | 80 | 30 |
| 23 | 147 | 89 | 80 | 30 |
| 25 | 147 | 89 | 80 | 30 |
| 28 | 148 | 90 | 81 | 35 |
| 30 | 149 | 91 | 50 | 36 |
| 32 | 150 | 92 | 82 | 37 |
| 35 | 151 | 93 | 83 | 38 |
| 78 | 151 | 93 | 83 | 38 |
| 36 | 152 | 94 | 84 | 25 |
| 38 | 153 | 95 | 85 | 39 |
| 39 | 154 | 96 | 12 | 40 |
| 4 | 155 | 97 | 86 | 30 |
| 41 | 156 | 98 | 87 | 41 |

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| 44 | 156 | 98 | 87 | 41 |
| 42 | 157 | 99 | 88 | 25 |
| 45 | 158 | 100 | 84 | 42 |
| 51 | 159 | 101 | 89 | 43 |
| 52 | 159 | 101 | 89 | 43 |
| 53 | 160 | 102 | 90 | 44 |
| 54 | 161 | 103 | 7 | 45 |
| 56 | 162 | 104 | 74 | 30 |
| 57 | 163 | 105 | 10 | 46 |
| 58 | 163 | 105 | 10 | 46 |
| 59 | 163 | 105 | 10 | 46 |
| 60 | 163 | 105 | 10 | 46 |
| 62 | 164 | 106 | 91 | 25 |

| Strain | ST | *recA* | *mdh* | *ompK* |
|---|---|---|---|---|
| 63 | 165 | 107 | 92 | 46 |
| 91 | 165 | 107 | 92 | 46 |
| 95 | 165 | 107 | 92 | 46 |
| 67 | 166 | 108 | 94 | 47 |
| 68 | 166 | 108 | 94 | 47 |
| 7 | 167 | 110 | 96 | 48 |
| 76 | 168 | 111 | 98 | 50 |
| 80 | 169 | 112 | 58 | 52 |
| 82 | 170 | 113 | 99 | 17 |
| 83 | 170 | 113 | 99 | 17 |
| 84 | 170 | 113 | 99 | 17 |
| 93 | 171 | 114 | 100 | 53 |

nt-ST & aa-ST cross reference table

| Strain | nt-ST | aa-ST |
|---|---|---|
| 1 | 111 | 64 |
| 2 | 111 | 64 |
| 4 | 155 | 112 |
| 5 | 111 | 64 |
| 7 | 167 | 67 |
| 12 | 119 | 84 |
| 13 | 117 | 103 |
| 14 | 124 | 105 |
| 15 | 117 | 103 |
| 16 | 127 | 65 |
| 17 | 127 | 65 |
| 18 | 134 | 68 |
| 19 | 127 | 65 |
| 20 | 145 | 96 |
| 21 | 147 | 112 |
| 23 | 147 | 112 |
| 25 | 147 | 112 |
| 28 | 148 | 109 |
| 30 | 149 | 70 |
| 32 | 150 | 110 |
| 33 | 122 | 86 |
| 34 | 122 | 86 |
| 35 | 151 | 75 |
| 36 | 152 | 69 |
| 38 | 153 | 79 |
| 39 | 154 | 71 |
| 41 | 156 | 72 |
| 42 | 157 | 69 |
| 43 | 172 | 86 |
| 44 | 156 | 72 |
| 45 | 158 | 75 |
| 46 | 61 | 86 |
| 47 | 135 | 70 |
| 51 | 159 | 80 |
| 52 | 159 | 81 |
| 53 | 160 | 73 |
| 54 | 161 | 74 |
| 56 | 162 | 112 |
| 57 | 163 | 75 |
| 58 | 163 | 75 |
| 59 | 163 | 75 |
| 60 | 163 | 75 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| 61 | 112 | 70 |
| 62 | 164 | 69 |
| 63 | 165 | 75 |
| 64 | 120 | 86 |
| 66 | 82 | 64 |
| 67 | 166 | 89 |
| 68 | 166 | 89 |
| 72 | 97 | 88 |
| 76 | 168 | 112 |
| 77 | 63 | 86 |
| 78 | 151 | 75 |
| 80 | 169 | 105 |
| 82 | 170 | 65 |
| 83 | 170 | 65 |
| 84 | 170 | 65 |
| 91 | 165 | 75 |
| 93 | 171 | 108 |
| 95 | 165 | 75 |
| 96 | 146 | 97 |
| 98 | 122 | 86 |
| 100 | 113 | 82 |
| 101 | 113 | 82 |
| 103 | 113 | 82 |
| 105 | 58 | 83 |
| 106 | 114 | 91 |
| 108 | 115 | 93 |
| 109 | 116 | 102 |
| 112 | 118 | 104 |
| 113 | 114 | 92 |
| 130 | 113 | 82 |
| 133 | 121 | 85 |
| 134 | 61 | 86 |
| 137 | 123 | 66 |
| 141 | 113 | 82 |
| 142 | 81 | 77 |
| 143 | 81 | 77 |
| 145 | 113 | 82 |
| 146 | 113 | 82 |
| 149 | 62 | 86 |
| 154 | 125 | 106 |
| 158 | 126 | 98 |
| 161 | 128 | 94 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| 162 | 129 | 99 |
| 163 | 130 | 78 |
| 164 | 130 | 78 |
| 166 | 130 | 78 |
| 172 | 131 | 99 |
| 176 | 132 | 107 |
| 178 | 133 | 100 |
| 180 | 133 | 100 |
| 186 | 136 | 69 |
| 187 | 137 | 101 |
| 188 | 138 | 86 |
| 189 | 140 | 64 |
| 190 | 142 | 95 |
| 191 | 143 | 112 |
| 192 | 141 | 64 |
| 194 | 139 | 86 |
| 195 | 118 | 104 |
| 199 | 144 | 90 |
| 200 | 64 | 87 |
| 202 | 146 | 97 |
| 206 | 61 | 86 |
| S1A65i | 60 | 48 |
| S1A65iii | 15 | 14 |
| S1A65ix | 14 | 19 |
| S1A65v | 3 | 14 |
| S1A65vi | 17 | 14 |
| S1A65vii | 14 | 19 |
| S1A65x | 14 | 19 |
| S1A66i | 5 | 18 |
| S1A66ii | 3 | 14 |
| S1A66iii | 21 | 14 |
| S1A66iv | 19 | 14 |
| S1C01i | 91 | 14 |
| S1C01ii | 19 | 14 |
| S1C01iii | 18 | 19 |
| S1C01iv | 91 | 14 |
| S1L01ii | 3 | 14 |
| S1L01iii | 89 | 60 |
| S1L01iv | 90 | 61 |
| S1L01vi | 19 | 14 |
| S1L02i | 13 | 14 |
| S1L02iii | 3 | 14 |
| S1L02v | 109 | 63 |
| S1M01i | 88 | 14 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S1M01ii | 3 | 14 |
| S1M01iii | 3 | 14 |
| S1M01iv | 13 | 14 |
| S1M01v | 13 | 14 |
| S1M02i | 3 | 14 |
| S1M02iii | 3 | 14 |
| S1M02iv | 3 | 14 |
| S1M02v | 11 | 31 |
| S1P01i | 87 | 52 |
| S1P01ii | 87 | 52 |
| S1S18i | 13 | 14 |
| S1S18ii | 4 | 14 |
| S1S18iii | 3 | 14 |
| S1S18iv | 12 | 18 |
| S1S18v | 3 | 14 |
| S2aA02vi | 108 | 15 |
| S2aA02vii | 19 | 14 |
| S2aA02viii | 3 | 14 |
| S2aA03i | 51 | 8 |
| S2aA03ii | 50 | 8 |
| S2aA03vi | 42 | 46 |
| S2aA03vii | 43 | 5 |
| S2aA04i | 3 | 14 |
| S2aA04ii | 3 | 14 |
| S2aA04iii | 48 | 7 |
| S2aA04iv | 49 | 29 |
| S2aA04vi | 28 | 2 |
| S2aA04vii | 28 | 2 |
| S2aA04x | 3 | 14 |
| S2aA05ix | 67 | 10 |
| S2aA05vi | 10 | 10 |
| S2aA05x | 9 | 3 |
| S2aA06ix | 16 | 23 |
| S2aA06vi | 16 | 23 |
| S2aA06vii | 16 | 23 |
| S2aA06viii | 23 | 111 |
| S2aA06x | 16 | 23 |
| S2aA09ix | 1 | 1 |
| S2aA09vi | 2 | 13 |
| S2aA09vii | 3 | 14 |
| S2aA09viii | 13 | 14 |
| S2aA09x | 3 | 14 |
| S2aA10iii | 56 | 3 |
| S2aA11ix | 3 | 14 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S2aA13i | 52 | 48 |
| S2aA13ii | 13 | 14 |
| S2aA13iv | 54 | 26 |
| S2aA13v | 3 | 14 |
| S2aA14i | 3 | 14 |
| S2aA14iv | 3 | 14 |
| S2aA28i | 46 | 2 |
| S2aA28v | 3 | 14 |
| S2aA29ix | 34 | 3 |
| S2aA29vi | 35 | 16 |
| S2aA29viii | 37 | 38 |
| S2aA29x | 37 | 38 |
| S2aA30ix | 36 | 35 |
| S2aA30vi | 38 | 38 |
| S2aA30vii | 13 | 14 |
| S2aA30viii | 66 | 3 |
| S2aA30x | 68 | 3 |
| S2aA31i | 31 | 37 |
| S2aA31ii | 47 | 33 |
| S2aA31iii | 57 | 56 |
| S2aA32ix | 3 | 14 |
| S2aA32vi | 3 | 14 |
| S2aA32vii | 3 | 14 |
| S2aA32viii | 3 | 14 |
| S2aA32x | 23 | 111 |
| S2aA34viii | 30 | 36 |
| S2aA34x | 33 | 25 |
| S2aA67ap | 59 | 48 |
| S2aS01iii | 44 | 6 |
| S2bA41vii | 26 | 39 |
| S2bA41viii | 56 | 3 |
| S2bA41x | 69 | 22 |
| S2bA42i | 5 | 18 |
| S2bA42ii | 3 | 14 |
| S2bA42v | 6 | 19 |
| S2bA44viii | 22 | 35 |
| S2bA44x | 13 | 14 |
| S2bA46x | 74 | 9 |
| S2bA48iii | 78 | 12 |
| S2bA48iv | 39 | 43 |
| S2bA48v | 79 | 18 |
| S2bA48vi | 70 | 3 |
| S2bA48vii | 14 | 19 |
| S2bA48viii | 85 | 30 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S2bA48x | 55 | 27 |
| S2bA50vii | 13 | 14 |
| S2bA50viii | 17 | 14 |
| S2bA51iii | 75 | 14 |
| S2bA51iv | 6 | 19 |
| S2bA52vi | 3 | 14 |
| S2bA52viii | 73 | 58 |
| S2bA53vi | 70 | 3 |
| S2bA53vii | 3 | 14 |
| S2bA53viii | 65 | 11 |
| S2bA53x | 173 | 57 |
| S2bA57i | 80 | 17 |
| S2bA57iii | 83 | 24 |
| S2bA57vi | 13 | 14 |
| S2bA57vii | 3 | 14 |
| S2bA58ii | 84 | 59 |
| S2bA58iii | 13 | 14 |
| S2bA59ii | 25 | 34 |
| S2bA61v | 3 | 14 |
| S2bA62i | 76 | 24 |
| S2bA62ii | 70 | 3 |
| S2bA62iii | 53 | 14 |
| S2bA62iv | 70 | 3 |
| S2bA62vi | 29 | 3 |
| S2bA62viii | 13 | 14 |
| S2bA63ix | 71 | 4 |
| S2bA63vi | 72 | 8 |
| S2bA64vi | 13 | 14 |
| S2bA64vii | 3 | 14 |
| S2bA64x | 3 | 14 |
| S2bA69ap | 86 | 49 |
| S2bA70ap | 60 | 48 |
| S3A72ii | 92 | 62 |
| S3A73i | 40 | 40 |
| S3A75ii | 93 | 44 |
| S3A75v | 94 | 36 |
| S3A75vi | 94 | 22 |
| S3A77v | 32 | 38 |
| S3A77vii | 27 | 42 |
| S3A79i | 24 | 41 |
| S3L05i | 41 | 45 |
| S4aA80ii | 104 | 51 |
| S4aA80iii | 105 | 3 |
| S4aA80iv | 7 | 3 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S4aA81i | 106 | 19 |
| S4aA81iii | 3 | 32 |
| S4aA81v | 104 | 14 |
| S4aP04iii | 101 | 3 |
| S4aP08i | 102 | 54 |
| S4aP10ii | 103 | 54 |
| S4aP10iii | 110 | 53 |
| S4aP10iv | 3 | 52 |
| S4aP10v | 13 | 14 |
| S4aS05v | 95 | 14 |
| S4aS09i | 6 | 47 |
| S4aS09ii | 3 | 19 |
| S4aS09iv | 98 | 14 |

| Strain | nt-ST | aa-ST |
|---|---|---|
| S4aS09v | 99 | 54 |
| S4bA84ii | 8 | 20 |
| S4bA84iii | 107 | 21 |
| S4bA84v | 107 | 21 |
| S4bP14iv | 56 | 3 |
| S4bP14v | 45 | 6 |
| S4bP16ii | 3 | 14 |
| S4bP16v | 96 | 50 |
| S4bP18i | 3 | 14 |
| S4bP18iii | 13 | 14 |
| S4bS11i | 100 | 55 |
| S4bS14i | 77 | 28 |

# APPENDIX E

## SUPPLEMENTARY INFORMATION FOR CHAPTER SEVEN

Details of *S. aureus* strains used in the *agr* competition assays (n=45)

| Strain | *agr* | ST |
|---|---|---|
| OX2 | 1 | 16 |
| OX21 | 3 | 30 |
| OX42 | 1 | 16 |
| OX67 | 1 | 8 |
| OX137 | 3 | 30 |
| OX154 | 2 | 15 |
| OX162 | 3 | 1 |
| OX208 | 1 | 22 |
| OX279 | 3 | 30 |
| OX303 | 2 | 12 |
| OX316 | 2 | 16 |
| OX357 | 2 | 15 |
| OX390 | 3 | 30 |
| OX410 | 2 | 15 |
| OX434 | 1 | 8 |
| OX455 | 3 | 30 |
| OX466 | 2 | 5 |
| OX481 | 1 | 45 |
| OX525 | 2 | 12 |
| OX563 | 3 | 37 |
| OX640 | 1 | 22 |
| OX724 | 1 | 22 |
| OX811 | 2 | 5 |
| OX858 | 3 | 30 |
| OX952 | 3 | 39 |
| OX3003 | 1 | 16 |
| OX3009 | 2 | 15 |
| OX3016 | 1 | 25 |
| OX3022 | 2 | 15 |
| OX3029 | 1 | 16 |
| OX3035 | 1 | 25 |
| OX3041 | 1 | 45 |
| OX3048 | 2 | 15 |
| OX3054 | 1 | 45 |
| OX3077 | 1 | 16 |
| OX3083 | 3 | 1 |
| OX3090 | 1 | 45 |
| OX3103 | 3 | 30 |
| OX3109 | 2 | 15 |
| OX3116 | 1 | 16 |
| OX3123 | 3 | 51 |
| OX3129 | 3 | 30 |
| OX3137 | 1 | 8 |
| OX3151 | 1 | 1 |
| OX3158 | 2 | 15 |
| OX3172 | 1 | 8 |
| OX3177 | 1 | 16 |
| OX3183 | 1 | 25 |

*S. aureus* primers

***arcC***
For 5' TTG ATT CAC CAG CGC GTA TTG TC 3'
Rev 5' AGG TAT CTG CTT CAA TCA GCG 3'

***aroE***
For 5' ATC GGA AAT CCT ATT TCA CAT TC 3'
Rev 5' GGT GTT GTA TTA ATA ACG ATA TC 3'

***glpF***
For 5' CTA GGA ACT GCA ATC TTA ATC C 3'
Rev 5' TGG TAA AAT CGC ATG TCC AAT TC 3'

***gmk***
For 5' ATC GTT TTA TCG GGA CCA TC 3'
Rev 5' TCA TTA ACT ACA ACG TAA TCG TA 3'

***pta***
For 5' GTT AAA ATC GTA TTA CCT GAA GG 3'
Rev 5' GAC CCT TTT GTT GAA AAG CTT AA 3'

***tpi***
For 5' TCG TTC ATT CTG AAC GTC GTG AA 3'
Rev 5' TTT GCA CCT TCT AAC AAT TGT AC 3'

***yqiL***
For 5' CAG CAT ACA GGA CAC CTA TTG GC 3'
Rev 5' CGT TGA GGA ATC GAT ACT GGA AC 3'

JOURNAL OF BACTERIOLOGY, Nov. 2006, p. 7686–7688 Vol. 188, No. 21
0021-9193/06/$08.00+0 doi:10.1128/JB.00700-06


# NOTES

## Agr Interference between Clinical *Staphylococcus aureus* Strains in an Insect Model of Virulence[▽]†

Vicki Fleming,[1] Ed Feil,[1] Andrew K. Sewell,[2] Nicholas Day,[3] Angus Buckling,[4] and Ruth C. Massey[4]*

*Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom[1]; Peter Medawar Building, Nuffield Department of Medicine, University of Oxford, Oxford OX1 3SY, United Kingdom[2]; Faculty of Tropical Medicine, Mahidol University, 420/6 Rajvithi Road, Bangkok 10400, Thailand[3]; and Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom[4]*

Received 17 May 2006/Accepted 16 August 2006

**Repression of virulence by *Staphylococcus aureus* strains from different Agr groups has been demonstrated in vitro and is proposed as a means of competitive interference. Here, using the insect *Manduca sexta*, we show for the first time that this interference also occurs in vivo within a mixed population.**

*Staphylococcus aureus* is a major human pathogen, and with strains that are resistant to all antibiotics in clinical usage (e.g., methicillin- and vancomycin-resistant *Staphylococcus aureus*) emerging worldwide, it is imperative that further studies investigating potential routes of virulence management are pursued. One approach is to identify genetic factors associated with *S. aureus*' competitive ability in vivo. The development of insect models of virulence for human pathogens provides financial, administrative, and ethical advantages over the use of mammalian models. For *S. aureus*, three have been developed to date: the silkworm *Bombyx mori* (6), the fruit fly *Drosophila melanogaster* (11), and the roundworm *Caenorhabditis elegans* (15). However, these models are limited by a single, major factor in that they cannot be incubated at 37°C, the physiologically relevant temperature for human pathogens. In this study, we used *Manduca sexta*, the tobacco hornworm, for modeling *S. aureus* infections, as it can be incubated at 37°C, is cheap to produce in large numbers and easy to handle, has a well-studied physiology, immunology, and anatomy, and has a large size, which facilitates accurate inoculation, straightforward dissections, and a simple index of virulence through weight loss and mortality.

The *S. aureus* strain collection used in this study has been described and characterized in detail elsewhere (2, 9, 13, 14). The strains were grown overnight at 37°C in 5 ml of brain heart infusion broth. The hemocoels of fifth-instar *M. sexta* larvae were injected with 10 μl of phosphate-buffered saline containing washed bacterial cells at a density of $10^4$ CFU per insect. Photographs of fifth-instar *M. sexta* larvae before inoculation and after 3 days of incubation at 37°C can be seen in Fig. 1a. Uninfected larvae grow up to 8 cm in length and gain up to 5 g during this time; in contrast, larvae infected with *S. aureus* did not grow and underwent mass necrosis followed by death. The growth dynamics of 13 diverse *S. aureus* strains (experiments performed in triplicate) (see Table S1 in the supplemental material) in *M. sexta* and the corresponding changes in weight of the infected insects relative to an uninfected control can be seen in Fig. 1b. Note that no reproducible differences in bacterial growth or insect mortality/growth were observed across these 13 strains.

The aim of this study was to identify genetic factors that correlated with *S. aureus*' competitive ability in vivo. A random subset of strains from the collection described above (45 strains) was competed against an individual, marked *S. aureus* strain to measure the strains' relative fitness. We then looked for associations between relative strain fitness and specific genetic factors. A naturally tetracycline-resistant strain was identified from the collection and is hereafter termed the marked strain. The hemocoels of fifth-instar *M. sexta* larvae were injected with 10 μl of medium containing equal numbers of the two competing strains at a density of $10^4$ CFU per insect. The starting densities were measured by plating on Trypticase soy agar with or without tetracycline (2 μg/ml) prior to inoculation. The insects were then incubated at 37°C for 3 days and homogenized, and the homogenate was serially diluted and plated on Trypticase soy agar with or without tetracycline. Any insects that died before this final time point were discarded. After 24-h incubation at 37°C, the marked (tetracycline-resistant) and tester (tetracycline-sensitive) colonies were enumerated. Relative fitness ($m$) was calculated from the ratio of the marked/tester strain Malthusian parameters [$m = \ln(N_f/N_o)$, where $N_f$ and $N_o$ are the final and starting densities, respectively] (7). The competition experiments were performed in duplicate and the results averaged.

We fitted a variety of factors (e.g., presence of specific genes, sequence type, and resistance to different antibiotics) to a general linear model (GLM) factor simultaneously, and the only factor that contributed to increased relative fitness in vivo was Agr type ($P < 0.01$) (see Table S1 in the supplemental

* Corresponding author. Mailing address: Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom. Phone: 44-1865271100. Fax: 44-1865310447. E-mail: ruth.massey@zoo.ox.ac.uk.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 25 August 2006.

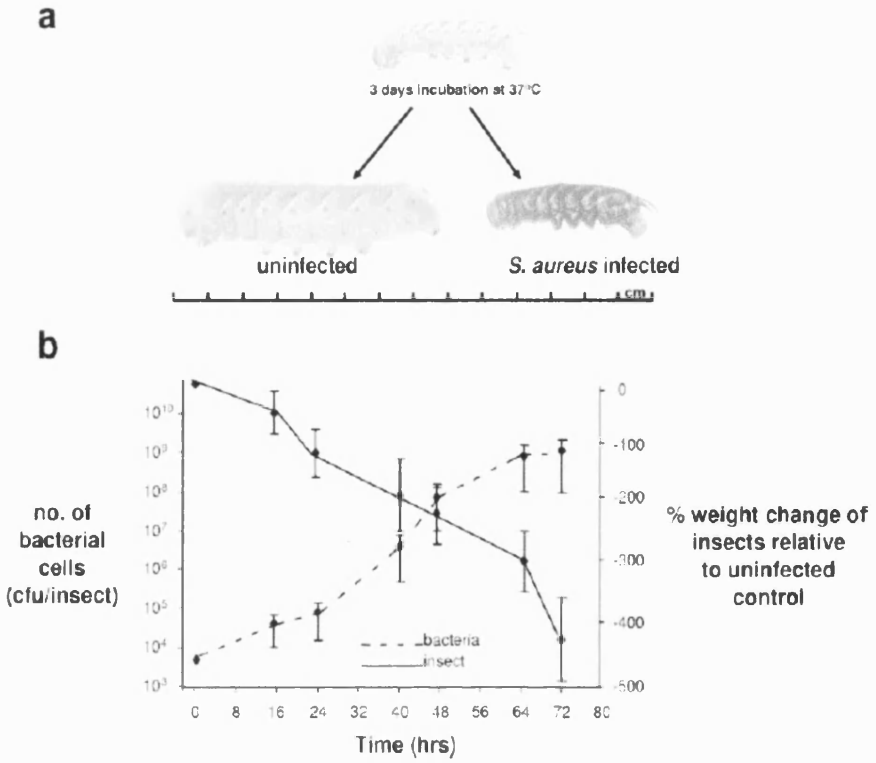



FIG. 1. *S. aureus* infection of *Manduca sexta*. (a) Photographs of *M. sexta* larvae before inoculation and after 3 days of incubation at 37°C after infection with $10^4$ *S. aureus* cells, alongside an uninfected control maintained under the same conditions. A scale (measuring in centimeters) is provided for comparison. (b) A graph showing the growth of 13 diverse *S. aureus* strains (see Table S1 in the supplemental material) within the insects and the corresponding weight changes of the insects relative to an uninfected control. The left-hand *y* axis shows the number of CFU of *S. aureus* per insect. The right-hand *y* axis shows the percent weight change of infected *M. sexta* larvae relative to the weight of the uninfected control. The error bars represent the standard deviations from the means.

TABLE 1. Agr interference occurs between competing *S. aureus* strains during infection[a]

| Marked-strain type | Relative fitness of competitor strain (mean ± SEM) | | |
|---|---|---|---|
| | Agr1 | Agr2 | Agr3 |
| Agr1 | 1 | 1.18 ± 0.12 | 1.31 ± 0.17 |
| Agr2 | 1.25 ± 0.06 | 1 | 0.84 ± 0.06 |
| Agr3 | 1.12 ± 0.17 | 1.06 ± 0.14 | 1 |

[a] The mean relative fitness values (± the standard error of the mean) of the different Agr strains in competitions against marked strains of Agr types 1, 2, and 3 are shown. For the purpose of making the data easier to interpret, the data were transformed such that when the tester strains were competed against the marked strain of the same Agr type, the mean fitness was converted to 1. This was achieved by multiplying the mean fitness values and the standard errors by 2.03 for the competitions against the marked Agr1 strains, by 4.2 for the competitions against the marked Agr2 strains, and by 3.42 for the competitions against the marked Agr3 strains. Data were analyzed using the general linear mixed model, with strain (treated as a random factor) nested in Agr type, the Agr type of the competitor, and the interaction between strain Agr type and competitor Agr type fitted. Relative fitness was first $\log_{10}$ transformed to meet GLM assumptions. There were significant differences between the marked competitors ($F_{2,82} = 58.53$, $P < 0.001$) and, crucially, a significant interaction between the Agr types of the competing strains ($F_{4,82} = 3.47$, $P = 0.01$). This interaction reveals that the outcome of competition between *S. aureus* strains is dependent on both competitors' Agr types.

material). The Agr quorum-sensing system regulates the expression of many genes in a cell density-dependent manner (12, 16). Recent work has established the existence of multiple (four) Agr types in *S. aureus* (3, 4, 10) and at least 20 others in related species (1) in which interference between different Agr types has been observed (4). The marked strain was of Agr type 2 (Agr2), and we found that the fitness of competing strains went in the order Agr1, Agr2, Agr3 (Table 1). This suggests either that certain Agr types are more successful at growing in this environment or, given the nature of the experiment (i.e., two strains competing within the same insect), that Agr interference between the competitor and marked strains is occurring within the insect.

If interference occurred between specific Agr types, we would expect to see the order of fitness of the different Agr types to vary depending on the Agr type they were competing against. Note that the large diversity of genetic backgrounds within this collection should, if anything, obscure any such interactions. By contrast, if the Agr type was simply correlated with different absolute competitive abilities, the order of fitness of Agr types should remain the same, regardless of the marked strain. To address this, we repeated the competitions against marked strains of Agr types 1 and 3. A GLM was used as

described above for these experiments. We found that relative fitness within the insect was dependent upon the combination of the Agr types of the competing strains, where the ascending order of fitness against marked strains 2 and 3 was 3-2-1, but that against marked strain 1 was 1-2-3 (Table 1), suggesting that Agr types interfere with each other. It is, however, possible that other factors in linkage disequilibrium with the Agr type and not evident through genetic analysis of the strain collection could also be responsible for the observed interference.

Interestingly, no effect of Agr type was detectable when competitions were performed in brain heart infusion broth. That no effect of Agr on competitive ability was observed under these conditions suggests that factors affected by Agr, such as toxicity and adhesiveness, contribute only to fitness in vivo. This is consistent with the role of Agr activation during infection, in which we also found that an Agr mutant of *S. aureus* was significantly attenuated in *M. sexta* (strain PC6911 [8325-4 *agrΔ::tet*], a generous gift from Simon Foster). After 24 h, the larvae ($n = 20$) infected with the wild-type strains lost on average 0.49 g, whereas the larvae infected with the Agr mutant gained 0.53 g (by two-sample $t$ test, $t = 5.53$; $P < 0.0001$). After 72 h, 100% of the larvae infected with the wild-type strain were dead, whereas only 20% of those infected with the Agr mutant strain were dead ($P < 0.001$).

Previous work has looked at the inhibitory effect of supernatant containing autoinducing peptides or purified/synthesized autoinducing peptides on Agr activation of in vitro populations of *S. aureus* (3, 4) and examined how the normal flora of nasal cavities changes over time (5, 8). This study advances these works by demonstrating that Agr interference can occur within a mixed population in vivo and that the competitive ability of a given Agr type depends on what Agr type it is competing against.

## REFERENCES

1. **Dufour, P., S. Jarraud, F. Vandenesch, T. Greenland, R. P. Novick, M. Bes, J. Etienne, and G. Lina.** 2002. High genetic variability of the *agr* locus in *Staphylococcus* species. J. Bacteriol. 184:1180–1186.
2. **Fell, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. Day.** 2003. How clonal is *Staphylococcus aureus*? J. Bacteriol. 185:3307–3316.
3. **Jarraud, S., G. J. Lyon, A. M. Figueiredo, L. Gerard, F. Vandenesch, J. Etienne, T. W. Muir, and R. P. Novick.** 2000. Exfoliatin-producing strains define a fourth *agr* specificity group in *Staphylococcus aureus*. J. Bacteriol. 182:6517–6522.
4. **Ji, G., R. Beavis, and R. P. Novick.** 1997. Bacterial interference caused by autoinducing peptide variants. Science 276:2027–2030.
5. **Kahl, B. C., K. Becker, A. W. Friedrich, J. Clasen, B. Sinha, C. Von Eiff, and G. Peters.** 2003. *agr*-dependent bacterial interference has no impact on long-term colonization of *Staphylococcus aureus* during persistent airway infection of cystic fibrosis patients. J. Clin. Microbiol. 41:5199–5201.
6. **Kaito, C., N. Akimitsu, H. Watanabe, and K. Sekimizu.** 2002. Silkworm larvae as an animal model of bacterial infection pathogenic to humans. Microb. Pathog. 32:183–190.
7. **Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler.** 1991. Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315–1341.
8. **Lina, G., F. Bonttte, A. Tristan, M. Bes, J. Etienne, and F. Vandenesch.** 2003. Bacterial competition for human nasal cavity colonization: role of staphylococcal *agr* alleles. Appl. Environ. Microbiol. 69:18–23.
9. **Lindsay, J. A., C. E. Moore, N. P. Day, S. J. Peacock, A. A. Witney, R. A. Stabler, S. E. Husain, P. D. Butcher, and J. Hinds.** 2006. Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. J. Bacteriol. 188: 669–676.
10. **Lyon, G. J., and R. P. Novick.** 2004. Peptide signaling in *Staphylococcus aureus* and other gram-positive bacteria. Peptides 25:1389–1403.
11. **Needham, A. J., M. Kibart, H. Crossley, P. W. Ingham, and S. J. Foster.** 2004. *Drosophila melanogaster* as a model host for *Staphylococcus aureus* infection. Microbiology 150:2347–2355.
12. **Otto, M.** 2004. Quorum-sensing control in staphylococci—a target for antimicrobial drug therapy? FEMS Microbiol. Lett. 241:135–141.
13. **Peacock, S. J., C. E. Moore, A. Justice, M. Kantzanou, L. Story, K. Mackie, G. O'Neill, and N. P. Day.** 2002. Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus*. Infect. Immun. 70: 4987–4996.
14. **Roche, F. M., R. Massey, S. J. Peacock, N. P. Day, L. Visai, P. Speziale, A. Lam, M. Pallen, and T. J. Foster.** 2003. Characterization of novel LPXTG-containing proteins of *Staphylococcus aureus* identified from genome sequences. Microbiology 149:643–654.
15. **Sifri, C. D., J. Begun, F. M. Ausubel, and S. B. Calderwood.** 2003. *Caenorhabditis elegans* as a model host for *Staphylococcus aureus* pathogenesis. Infect. Immun. 71:2208–2217.
16. **Winzer, K., and P. Williams.** 2001. Quorum sensing and the regulation of virulence gene expression in pathogenic bacteria. Int. J. Med. Microbiol. 291:131–143.